Developmental Constraints Do Not Influence Long-Term Phenotypic Evolution of Marsupial Forelimbs as Revealed by Interspecific Disparity and Integration Patterns

Alberto Martín-Serra[1,2,*] and Roger B. J. Benson[1]

1. Department of Earth Sciences, University of Oxford, South Parks Road, OX1 3AN Oxford, United Kingdom;   2. Departamento de Ecología y Geología, Facultad de Ciencias, Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain

abstract: Marsupials show a smaller range of forelimb ecomorphologies than placental mammals, and it is hypothesized that this results from macroevolutionary constraints imposed by the specialized reproductive biology of marsupials. Specifically, the accelerated development of the marsupial forelimb allows neonates to crawl to the mother's pouch but may constrain adult morphology. This hypothesis makes three main predictions: (i) that marsupial forelimbs should show less interspecific disparity than their hindlimbs, (ii) that morphological integration within the marsupial forelimb is stronger than integration between limbs, and (iii) that these patterns should be strongest in diprotodontians, which undergo the most rigorous crawls as neonates. We use a three-dimensional geometric morphometric data set of limb bones for 51 marsupial species to test these predictions. We find that (i) marsupial forelimbs and hindlimbs show similar disparities, (ii) no clear differences in integration exist either within or between limbs, and (iii) the same patterns occur in diprotodontians as in other marsupials, even correcting for lineage age. Therefore, there is currently little evidence that the developmental biology of marsupials has constrained their macroevolutionary patterns. It is possible that functional selection can overcome the effects of developmental constraint on macroevolutionary timescales. Our findings suggest that the role of developmental constraints in explaining the limited phenotypic variability of marsupials (compared with that of placentals) should be reconsidered.

Keywords: marsupials, forelimb, hindlimb, disparity, integration, constraint.

Introduction

Adaptation due to natural selection can be constrained by patterns of development and by evolutionary history.

* Corresponding author; email: alberto.martinserra@earth.ox.ac.uk.
  ORCIDs: Martín-Serra, https://orcid.org/0000-0003-0221-1103.

This is of central importance in the study of phenotypic evolution (e.g., Maynard-Smith et al. 1985; Wagner 1988; Wagner and Altenberg 1996; Futuyma 2010). Development ultimately provides the pool of available phenotypes for natural selection (Maynard-Smith et al. 1985), so developmental canalization and stability should tend to limit phenotypic variation between and within individuals, respectively (Hallgrímsson et al. 2002).

However, to what extent these developmental constraints have influenced phenotypic evolution on large timescales (i.e., macroevolution) remains controversial (McKitrick 1993; Richardson and Chipman 2003). Undoubtedly, development imposes major constraints on the potential of evolution to generate novel variation (Maynard-Smith et al. 1985; Galis et al. 2018). However, these constraints may be weaker than hypothesized, and even some apparently major constraints can be broken, as has been demonstrated for some clades (e.g., in echinoderm body plan evolution; Lowe and Wray 1997). This might be expected on long, macroevolutionary time spans. Phylogenetic comparative studies provide a means to assess which developmental patterns have been modified or overridden over long time spans and which have not. In particular, quantification of phenotypic disparity, rates of evolution, and the strength of integration provide insights into the underlying patterns and outcomes of evolutionary history (Wagner and Altenberg 1996; Goswami et al. 2014). If hypothesized constraints did influence the macroevolution of a group, then it is expected that this group could have lower disparity than groups of similar age and species richness, have slower rates of evolution, or exhibit stronger integration between body parts.

Phenotypic integration describes the structure of covariation among traits across the body or within body

regions (e.g., the skull or the forelimb). Integration is commonly hypothesized to result from developmental or functional constraints (Wagner and Altenberg 1996; Hallgrímsson et al. 2002) and can influence the direction and magnitude of evolutionary change (Goswami et al. 2014). However, as noted above, natural selection can also modify developmental patterns to fit with adaptive requirements (Wagner and Altenberg 1996). In other words, it can increase or decrease the degree of integration and the presence of modules according to the functional interactions of traits, providing one mechanism of breaking inherited patterns of constraint (Wagner and Altenberg 1996).

Marsupials provide a long-standing hypothesis of developmental constraints on phenotypic evolution. Strikingly, some of the specialized morphologies of placental mammals have no equivalent among marsupials, including the flippers of cetaceans, wings of bats, and cursorial forelimbs of some hoofed mammals and carnivorans. It has been suggested that this results from constraints on the evolution of marsupial forelimbs due to their specialized reproductive biology (e.g., Sears 2004). The prenatal period of gestation is short in marsupials, which lack the chorioallantoic placenta seen in placental mammals. Marsupial neonates are small, are altricial, and must perform an early crawl to reach their mother's pouch, where they complete their development by suckling (e.g., Lillegraven 1975; Tyndale-Biscoe and Renfree 1987; Gemmel et al. 2000). At this stage, forelimbs are well developed to function as climbing tools (they have digits capable of grasping and pulling the body), whereas hindlimbs are simple undifferentiated buds that hang passively (Lillegraven 1975; Tyndale-Biscoe and Renfree 1987; Gemmel et al. 2000; Sears 2004). It is widely hypothesized that the morphological disparity of marsupial forelimbs is constrained by this early requirement (Lillegraven 1975; Sears 2004; Goswami et al. 2009; Bennett and Goswami 2011; Kelly and Sears 2011; Garland et al. 2017). Under this hypothesis, early ontogenetic specialization for the neonatal crawl limits the range of possible adult forelimb morphologies and, as a consequence, the variation available for natural selection (Sears 2004). Furthermore, the offset in timing of the fore- and hindlimb ossification in marsupials may also have reduced the strength of integration between their adult forelimb and hindlimb morphologies (Bennett and Goswami 2011). This hypothesis provides an explanation for the low phenotypic and ecological variability of marsupials in comparison with placentals, which do not share such developmental constraints (Sears 2004; Goswami et al. 2009; Bennett and Goswami 2011; Kelly and Sears 2011; Garland et al. 2017). However, other factors might also explain this difference, such as historical biogeographic distributions of these groups. For example, Sánchez-Villagra (2013) hypoth-

esized that the differences between marsupials and placentals may have extrinsic causes: first, the Cretaceous/Paleogene extinction affected metatherians more severely than eutherians; and second, metatherians have primarily been restricted to the southern continents during the Cenozoic. These occupy much less terrestrial surface than northern ones, providing a narrower range of environments and opportunities for adaptation among marsupials compared with boreoeutherian placentals.

Three predictions of the hypothesis of developmental constraint on marsupial limb evolution for comparative (interspecific) data are (i) that the disparity of forelimb morphologies attained by marsupials during their evolutionary history is lower than that of their hindlimbs (Sears 2004), (ii) that morphological integration within the marsupial forelimb is stronger than that between the fore- and hindlimb (Goswami et al. 2009; Bennett and Goswami 2011; Kelly and Sears 2011; Garland et al. 2017), and (iii) that these patterns are strongest in diprotodontians, which undergo the most rigorous crawls as neonates because they have to climb from the urogenital sinus upward to the forward-facing pouch of their mother (Tyndale-Biscoe and Renfree 1987; Veitch et al. 2000; Sears 2004). To date, only the first prediction has been tested at an interspecific level (and not at the intraspecific level), and it was tested only using linear measurements of the scapula and pelvis (Sears 2004). To the best of our knowledge, the second prediction has been explored only at a static (e.g., intraspecific) level (Goswami et al. 2009; Bennett and Goswami 2011; Kelly and Sears 2011; Garland et al. 2017). Phylogenetic comparative (i.e., interspecific) analysis of these predictions using the main limb bones has not been conducted. Because of this, it remains uncertain whether the hypothesized developmental constraints on marsupial limb evolution have actually played out on macroevolutionary timescales.

We test these predictions using three-dimensional geometric morphometric data for the major limb bones (i.e., scapula, humerus, radius, and ulna for the forelimb and pelvis, femur, and tibia for the hindlimb) across a wide sample of marsupials (51 of 334 extant species), representing all major groups and ecomorphologies. This provides quantitative information about variation in marsupial limb morphologies and the extent to which they might be explained by development constraints rather than other factors.

Material and Methods

Material

We used a data set of 51 marsupial species collected from different institutions (table S1; tables S1–S6 are available online). These species were selected to cover as much of

the taxonomic and ecological diversity of marsupials as possible. For each one, we obtained three-dimensional models of the main limb bones: scapula, humerus, radius, ulna, pelvis, femur, and tibia from micro–computed tomography data and surface scans (for more detailed information, see the section "Supplementary Information 2—Methods" in the supplemental PDF, available online; scans and three-dimensional surfaces are available at https://www .morphosource.org/Detail/ProjectDetail/Show/project_id /428). We digitized a set of 62 landmarks and semilandmark series for these bones using the software Avizo (FEI Visualization Sciences Group 2015; fig. 1; supplemental PDF, "Supplementary Information 2—Methods"). Landmarks and semilandmarks were selected to capture the most important traits of these bones, such as tuberosities, that indicate muscle attachments or edges of articular surfaces (fig. 1; supplemental PDF, "Supplementary Information 2—Methods"). Adult male specimens were preferred when possible to avoid the effect of sexual dimorphism or ontogenetic changes in morphology. However, it was not always possible to select males due to the scarcity of complete specimens of many marsupial species in European and North American collections. All specimens present complete eruption of permanent dentition, indicating adult or near-adult age. We followed this criterion instead of complete fusion of growth plates given that some growth plates do not fuse completely in adults of many marsupial species (Washburn 1946; Geiger et al. 2014). Landmark data, pruned phylogeny, and scripts for this study are available in the Dryad Digital Repository (https://doi.org/10 .5061/dryad.900ng75; Martín-Serra and Benson 2020).

Preliminary Analyses

The three-dimensional coordinates of the landmarks and semilandmarks were imported to R version 3.4.1 (R Core Team 2017). For each bone, we performed a generalized Procrustes superimposition with these raw coordinates (Dryden and Mardia 1998) to remove the effects of size, translation, and orientation. Semilandmarks were allowed to slide along their tangent vectors to minimize bending energy differences from the mean shape (e.g., Gunz and Mitteroecker 2013). To do this, we used the function gpagen of the geomorph package (ver. 3.1.1; Adams et al. 2017). This procedure provides a data set with the shape coordinates (Procrustes coordinates) and centroid size (a proxy for bone size computed from the raw coordinates).

We then conducted a distance-based regression of the Procrustes coordinates for each bone to test the presence of a common interspecific (or evolutionary) allometric pattern (allometric regression). To do this, we used the function procD.pgls of the geomorph package (Collyer et al.

2015) with log-transformed centroid size as the independent variable and Procrustes coordinates as dependent variables (using 999 iterations for significance testing; Adams 2014). This analysis requires a phylogeny. We therefore pruned the molecular phylogeny published by Mitchell et al. (2014) so it only contained the species in our data set (for more detailed information, see the section "Supplementary Information 2—Methods" in the supplemental PDF). This phylogeny was built using mitochondrial and nuclear genes of extant marsupials and recently extinct ones (e.g., *Thylacinus cynocephalus*), and it includes branch lengths in millions of years (Mitchell at al. 2014). We also tested whether the three marsupial orders that are well represented in our data set (Dasyuromorphia [$N = 9$]; Didelphimorphia [$N = 11$]; Diprotodontia [$N = 23$]) have different allometric patterns by using the marsupial order name as a categorical variable. This variable was not significant in any cases, when used as either a covariate or an interaction term (see "Results"). We therefore used the residuals of the common allometric regression in subsequent analyses as an allometry-corrected form of the shape data and compared these to results from analyzing the original Procrustes coordinates.

Principal Component Analyses (PCAs) and Disparity

A series of PCAs was performed using the residuals obtained from the allometric regression and the plotTangentSpace function of geomorph (Adams et al. 2017). This function also provides the consensus shape and the three-dimensional coordinates of the extreme shapes of each PC. For visualization purposes, we used the geomorph functions mshape, warpRefMesh, and plotRefToTarget (Adams et al. 2017) to transform these coordinates into three-dimensional surfaces that represent the theoretical shape changes along PC axes.

We computed the morphological disparity of each bone using the function morphol.disparity of geomorph (Adams et al. 2017). This function calculates Procrustes variance, which is by definition the morphological disparity (Zelditch et al. 2012). We carried out this procedure using both Procrustes coordinates and the residuals of our allometric regression and using both our original samples ($N_{\text{Dasyuromorphia}} = 9$; $N_{\text{Didelphimorphia}} = 11$; $N_{\text{Diprotodontia}} = 23$) and subsampled data with equal species counts per order ($N = 9$; subsampled using a custom script in R). We also repeated this for each marsupial order, including a permutation test to assess the statistical significance of the disparity differences for each bone among groups (based on 999 iterations; Adams et al. 2017).

We repeated this procedure after dividing marsupial species into small and large body size categories. The
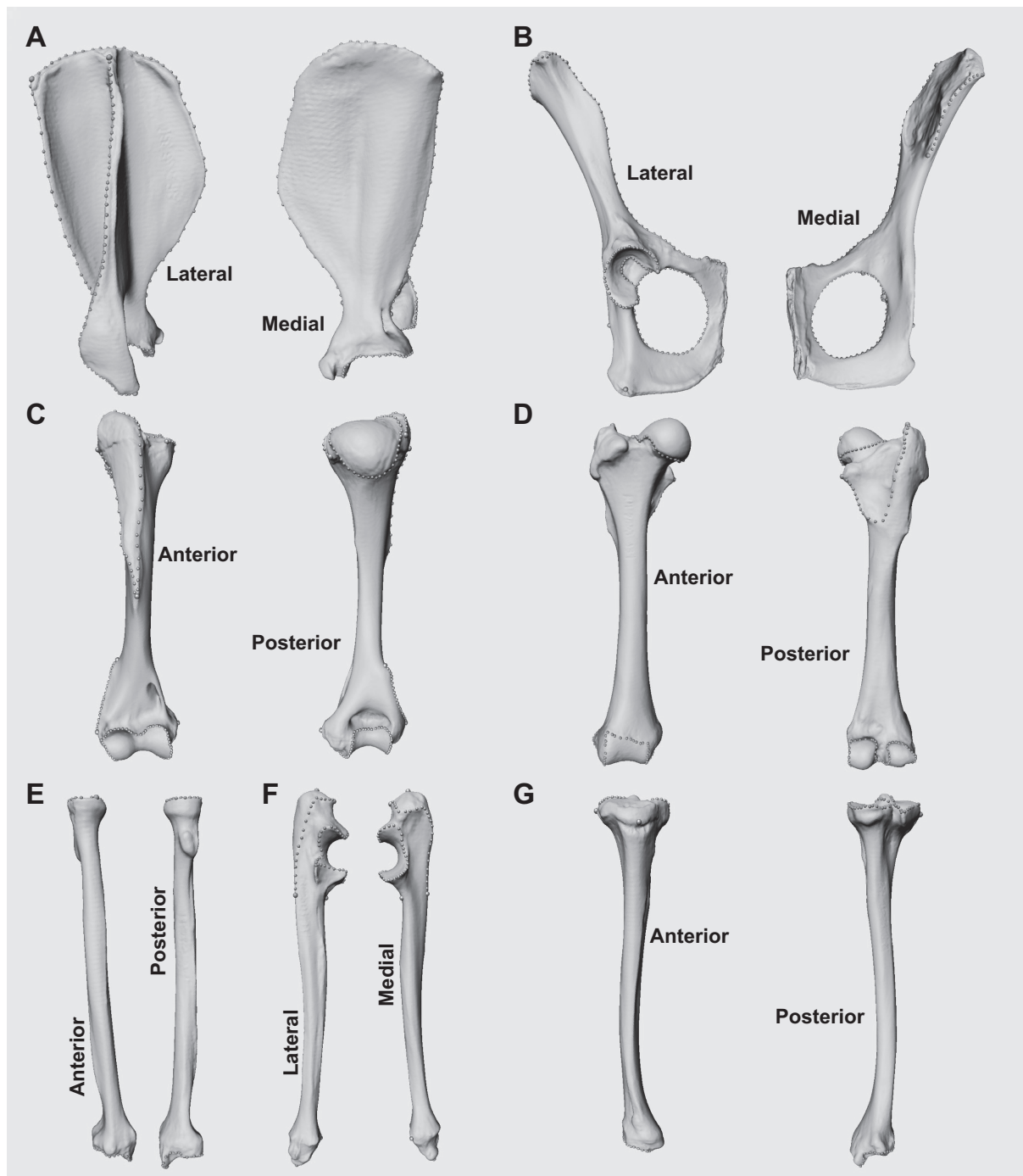
Figure 1: Landmarks and semilandmarks located on the limb bones analyzed. *A*, Scapula; *B*, half pelvis; *C*, humerus; *D*, femur; *E*, radius; *F*, ulna; *G*, tibia. All bones are right side and belong to the specimen *Sarcophilus harrisii* NHMUK 2003.331. Big gray spheres indicate landmarks, and small sphere rows indicate semilandmarks. For more details, see the section "Supplementary Information 2—Methods" in the supplemental PDF.

reasoning behind this is that acquiring a large body size can be accompanied by an increase in structural specialization for different ecologies (Stanley 1973). Accounting for allometric effects alone may not correct this effect if structural specializations of large-bodied species vary among ecologies. Diprotodontia reaches the largest body sizes of all marsupial orders (maximum body mass of ~45 kg, compared with ~23 kg in Dasyuromorphia and ~3 kg in

Didelphimorphia; Nowak 1999), and this might inflate their apparent disparity. To accommodate this possibility, we divided diprotodontians into two groups across a threshold body mass of 2 kg (for specific information and references, see table S1). This threshold was selected to so that both large ($N_{\text{large diprotodontians}} = 13$) and small ($N_{\text{small diprotodontians}} = 10$) categories contained similar number of species. Dasyuromorphia and Didelphimorphia were also divided into two categories (small and large) using same threshold. However, because these groups contain few species exceeding 2 kg in body mass (three in total across all nondiprotodontians in our data set), only the small-bodied species of these groups were analyzed. Permutation was used to assess the statistical significance of the differences in disparity among the resulting groups.

Evolutionary Rates

Differences in disparity among groups can result from differences in rates of evolution or from differences in the time available for clades to evolve (e.g., Collar et al. 2005; O'Meara et al. 2006; Sidlauskas 2008). The inferred age of the most basal node within Diprotodontia (53 Ma) is older than those of Didelphimorphia (38 Ma) or Dasyuromorphia (39 Ma) in the phylogeny of Mitchell et al. (2014), and this could cause Diprotodontia to be equally disparate or more disparate in the presence of evolutionary constraints. To address this, we used BayesTraits (ver. 3.0.1; http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.1/BayesTraitsV3.0.1.html) to estimate rates of multivariate evolution for the allometry-corrected residual shape data for each bone, broadly following the approach of Felice and Goswami (2018). To reduce dimensionality and associated problems (Adams and Collyer 2017), we performed PCA and subjected the scores of the PC axes that accounted for 95% of shape variance to an analysis of rates. We fit variable-rate Brownian motion models (Venditti et al. 2011; Baker et al. 2016; see also Eastman et al. 2011) using the independent contrasts method (Felsenstein 1973; Freckleton 2012) to increase speed of computation. Models were fit using a reversible-jump Markov chain Monte Carlo run for 10,000,000 generations, with the first 25% discarded as burn-in, and a sample period of 10,000 generations. This exceeds the number of iterations performed by similar previous analyses (e.g., Felice and Goswami 2018; 1,010,000 generations). Each analysis was run multiple times and yielded similar outcomes, suggesting that the pattern of rate variation returned by the analyses had converged. We compared the distributions of $\log_{10}$-transformed evolutionary rates within each group (Diprotodontia, Didelphimorphia, and Dasyuromorphia) to assess the differences among them.

Integration

To explore the strength of shape covariation among bones (a proxy for morphological integration), we selected a series of comparisons between pairs of bones according to two main criteria: bones that are anatomically connected within limbs (scapula-humerus, humerus-radius, humerus-ulna, and radius-ulna for the forelimb; pelvis-femur and femur-tibia for the hindlimb) and between-limb comparisons by limb portion (girdles: scapula-pelvis; stylopod: humerus-femur; zeugopod: radius-tibia and ulna-tibia). To do this, we carried out a series of two-block partial least squares analyses (Rohlf and Corti 2000) of the residuals from our allometric regression. We applied the geomorph functions two.b.pls (Adams and Collyer 2016) and phylo.integration (both with 999 iterations for evaluating statistical significance). The latter function explores morphological covariation along phylogenetic lineages assuming a Brownian model of evolution (Adams and Felice 2014; Adams and Collyer 2017). We also compared the differences in strength of integration among the comparisons using the Z value (Adams and Collyer 2016) computed by the function compare.pls. This function tests the significance of differences in integration coefficients by estimating the confidence interval of the difference of Z values at a given α. If this interval does not contain zero, the null hypothesis of absence of difference can be rejected (Adams and Collyer 2016). We repeated this procedure independently for Dasyuromorphia, Didelphimorphia, and Diprotodontia.

Results

PCAs and Disparity

Distance-based regression indicates that the most proximal appendicular bones (scapula, humerus, pelvis, and femur) show a significant common allometric pattern (table S2), but the distal ones (i.e., radius, ulna, and tibia) do not. The low explanatory power of all of these relationships indicates that the effect of common allometry is relatively small (table S2). This can result from (i) a general weak relationship between size and shape or (ii) large differences in the allometric patterns of individual groups, which are therefore overlooked when all groups are analyzed together. To assess this, we explored the possibility of differences in patterns of allometry between marsupial orders by using order names as interaction terms in the analysis. Order names are not significant for any bone, suggesting that all marsupial orders show the same allometric pattern (table S3). Therefore, size differences have a weak effect on marsupial limb bone morphologies. In spite of this, allometry may still result in apparently larger disparity for groups that span a wider range of body sizes (e.g., Diprotodontia). Therefore, all disparity computations have been done both before

and after accounting for common allometric effects (by analyzing the residuals of the distance-based regression of shape coordinates on size).

PCA results are shown and explained in the supplemental PDF ("Supplementary Information 3—Results," figs. S14 and S16; figs. S1–S26 are available online). For most limb bones, the marsupial mole (*Notoryctes typhlops*) is an outlier that drives a high portion of morphological variation; therefore, the PCAs were repeated without this species (figs. S15, S17). The main differences between the PCAs with and without the marsupial mole (for both fore- and hindlimb long bones) are a decrease in the range of variation of PC1, which accounts for changes in the robustness of each bone. This occurs because the marsupial mole has considerably more robust long bones than any other marsupial species. Only the scapula of the marsupial mole is not a clear outlier, resulting in almost no differences between the two scapula PCAs. Finally, for the pelvic girdle, the differences are more complex, as the marsupial mole is an outlier for PC1 and PC2, so when this species is removed both PCs change substantially.

The disparity values for the total sample and for Diprotodontia, Dasyuromorphia, and Didelphimorphia individually are shown in figure 2, and the major aspects of morphological variation among species are described using PCA (supplemental PDF, "Supplementary Information
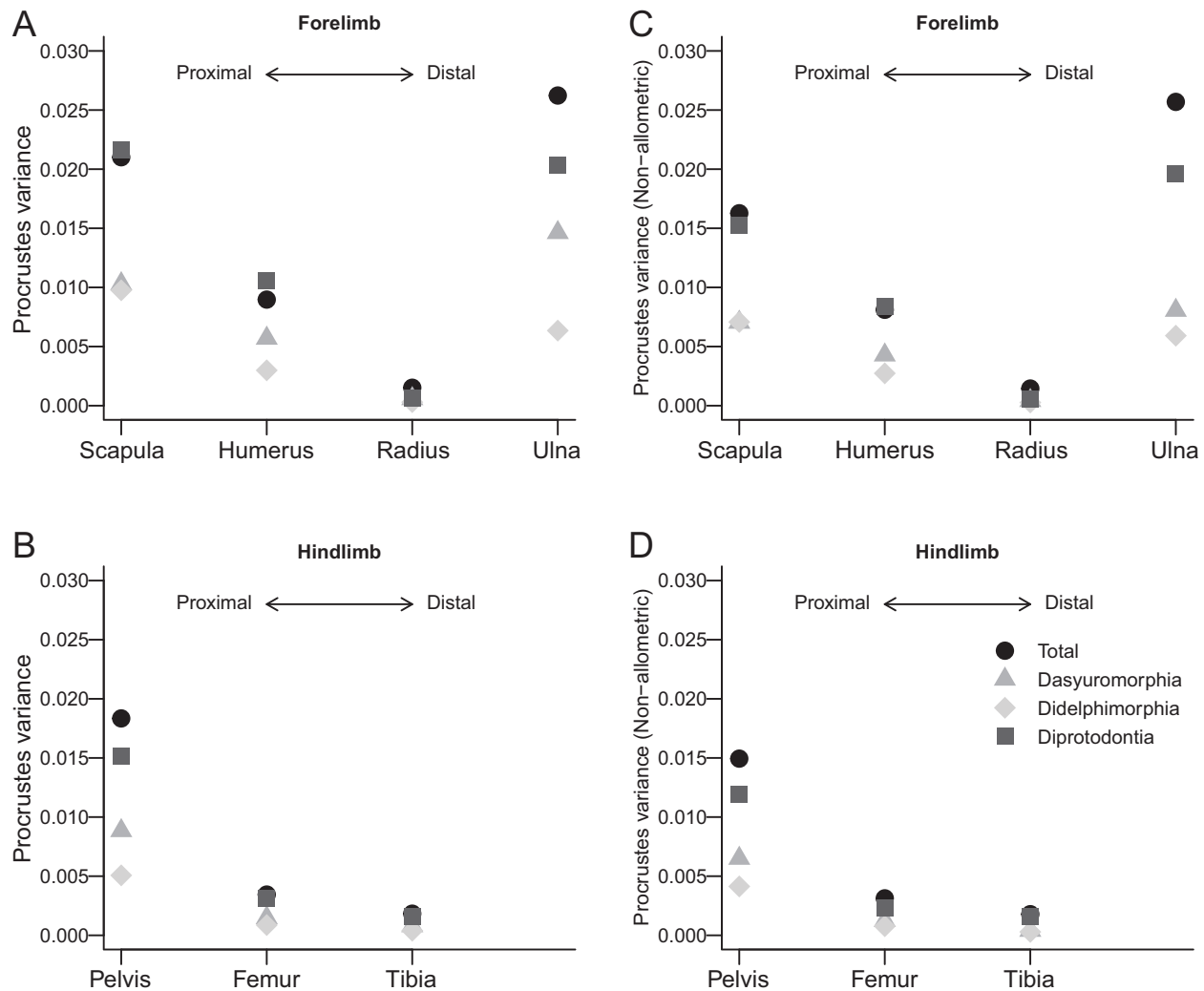

Figure 2: Morphological disparity of limb bones for marsupials and the three orders analyzed: Dasyuromorphia, Didelphimorphia, and Diprotodontia. *A*, Disparity values for the forelimb bones obtained from the Procrustes coordinates; *B*, disparity values for the hindlimb bones obtained from the Procrustes coordinates; *C*, disparity values for the forelimb bones obtained from the nonallometric residuals; *D*, disparity values for the hindlimb bones obtained from the nonallometric residuals.

3—Results"). For both data sets (corrected and not corrected for allometric changes), forelimb bones show disparities similar to those of the hindlimb among all taxonomic sample sets (fig. 2). Diprotodontians show greater disparity than the other orders (Dasyuromorphia, Didelphomorphia) for all bones, and these differences are statistically significant at the .05 level in 5 of 42 comparisons (11.9%; one asterisk in table 1) and at the .01 level in 9 of 42 (21.4%; two asterisks in table 1). Similar patterns are shown for median P values when groups are subsampled to equal sample sizes ($N = 9$; table S4), eliminating the possible bias due to different sample sizes. This does not result from the occurrence of large body size in some diprotodontians (e.g., vombatids, macropodids): diprotodontians still have disparities similar to those of other marsupials when small (average body weight from 11.5 g to 2 kg) and large (from 2 to 42.5 kg) diprotodontians are analyzed separately (fig. 3). Differences between the disparity of small diprotodontians and other small marsupials (Dasyuromorphia, Didelphomorphia) are nonsignificant (table 2).

Evolutionary Rates

Branch-specific distributions of evolutionary rates mapped to phylogenies and within groups (Diprotodontia, Dasyuromorphia, and Didelphimorphia) are displayed in figures S18–S24, and a summary of statistical comparisons is shown in table S5. Evolutionary rates of limb bone evolution in Diprotodontia are generally similar to those in Dasyuromorphia and Didelphimorphia, especially for the hindlimb (i.e., t-tests reject the hypothesis of differences in rates between clade; table S5). However, rates of humeral evolution are significantly lower in Dasyuromorphia than in Diprotodontia (table S5; figs. S18–S24), and Didelphimorphia show significantly lower rates of evolution of the scapula humerus and ulna than Diprotodontia (table S5).

Table 1: Results of the disparity comparisons between orders

	Dasyuromorphia	Didelphimorphia	Diprotodontia
Scapula:			
Dasyuromorphia	1	.976	.002**
Didelphimorphia	.905	1	.001**
Diprotodontia	.002**	.001**	1
Humerus:			
Dasyuromorphia	1	.62	.101
Didelphimorphia	.391	1	.018*
Diprotodontia	.082	.002**	1
Radius:			
Dasyuromorphia	1	.471	.345
Didelphimorphia	.174	1	.053
Diprotodontia	.929	.085	1
Ulna:			
Dasyuromorphia	1	.833	.127
Didelphimorphia	.472	1	.058
Diprotodontia	.666	.067	1
Pelvis:			
Dasyuromorphia	1	.5	.054
Didelphimorphia	.362	1	.002**
Diprotodontia	.058	.001**	1
Femur:			
Dasyuromorphia	1	.453	.028*
Didelphimorphia	.475	1	.001**
Diprotodontia	.02*	.002**	1
Tibia:			
Dasyuromorphia	1	.88	.051
Didelphimorphia	.597	1	.042*
Diprotodontia	.201	.041*	1

Note: P values obtained from the pairwise comparisons between orders are shown. Below-diagonal values use Procrustes-transformed shape data, and above-diagonal values use shape residuals after accounting for allometry.

* Statistically significant at a threshold of .05.

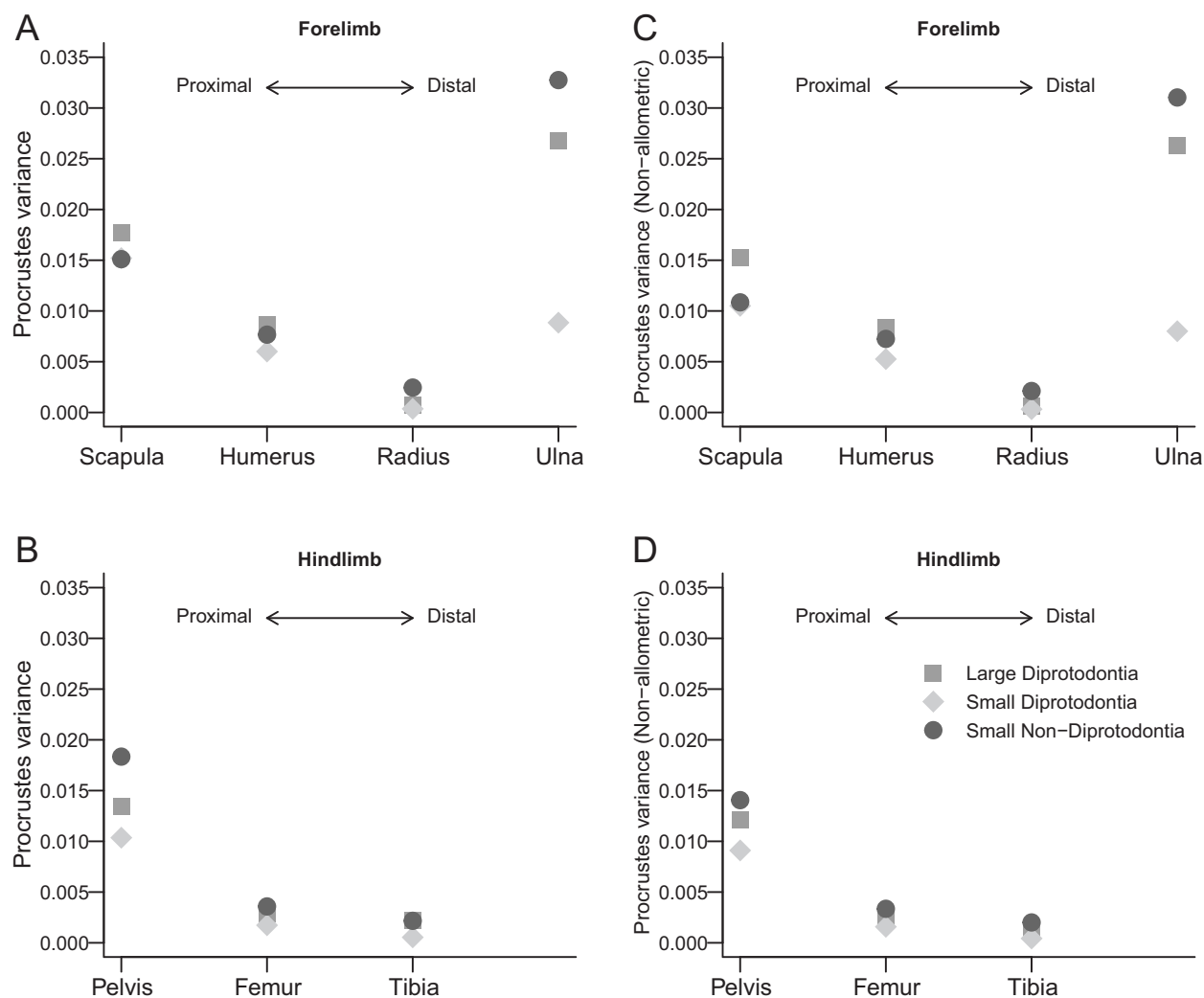** Statistically significant at a threshold of .01.

Figure 3: Morphological disparity of limb bones for small and large diprotodontians and small nondiprotodontian marsupials. *A*, Disparity values for the forelimb bones obtained from the Procrustes coordinates; *B*, disparity values for the hindlimb bones obtained from the Procrustes coordinates; *C*, disparity values for the forelimb bones obtained from the nonallometric residuals; *D*, disparity values for the hindlimb bones obtained from the nonallometric residuals.

Both groups also show lower rates of femoral evolution than Diprotodontia.

Integration

The results of two-block partial least squares indicate that the strength of integration (Z values) appears greater when ignoring the problem of phylogenetic autocorrelation than when accounting for it (figs. 4, S25; Adams and Felice 2014). Nevertheless, both sets of analyses show the same general pattern. The lowest integration values are associated with scapula-humerus (2.01 for phylogenetic integration) and scapula-pelvis (2.22) comparisons. In contrast, the greatest values are associated with radius-ulna and radius-tibia com-

parisons. Therefore, strength of integration seems to increase from low values among proximal bones (girdles) to higher values among distal bones (zeugopod). This applies for comparisons both within and between the fore- and hindlimbs.

The significance tests of the differences of strength of integration show that within-forelimb integration is not significantly greater than within-hindlimb or between-limb integration (tables 3, S6). In contrast, many of the statistically significant differences are associated with stronger integration between limbs; for instance, radius-tibia integration (5.7) is stronger than scapula-humerus and humerus-ulna integration (table 3). Most of the cases in which within-forelimb values are greater than within-hindlimb or

Table 2: Results of the disparity comparisons between size groups of diprotodontians and nondiprotodontians

	Large diprotodontians	Small diprotodontians	Small nondiprotodontians
Scapula:			
Large diprotodontians	1	.153	.093
Small diprotodontians	.489	1	.903
Small nondiprotodontians	.348	.967	1
Humerus:			
Large diprotodontians	1	.473	.778
Small diprotodontians	.513	1	.59
Small nondiprotodontians	.789	.646	1
Radius:			
Large diprotodontians	1	.652	.28
Small diprotodontians	.572	1	.225
Small nondiprotodontians	.501	.239	1
Ulna:			
Large diprotodontians	1	.489	.848
Small diprotodontians	.459	1	.307
Small nondiprotodontians	.804	.252	1
Pelvis:			
Large diprotodontians	1	.625	.769
Small diprotodontians	.714	1	.44
Small nondiprotodontians	.514	.281	1
Femur:			
Large diprotodontians	1	.571	.576
Small diprotodontians	.515	1	.271
Small nondiprotodontians	.707	.281	1
Tibia:			
Large diprotodontians	1	.503	.747
Small diprotodontians	.48	1	.245
Small nondiprotodontians	.999	.366	1

Note: P values obtained from the pairwise comparisons between size categories are shown. Below-diagonal values use Procrustes-transformed shape data, and above-diagonal values use shape residuals after accounting for allometry.

between-limbs values are associated with the strong integration between the radius and the ulna (tables 3, S6).

The pattern of integration for diprotodontians is very similar to that for the total sample: it shows lower values of integration when phylogeny is taken into account and a proximodistal gradient of increasing integration among limb bones (fig. 4, *bottom right*). However, dasyuromorphs and didelphimorphs display a different pattern (fig. 4, *top right and bottom left*), as the strengths of integration do not decrease when taking phylogeny into account and the proximodistal gradients of integration are less conspicuous (figs. 4, S25). A similar result occurs when comparing small and large diprotodontians individually with small nondiprotodontian marsupials (fig. S26).

In comparisons of diprotodontians, dasyuromorphs, and didelphimorphs, differences in the strength of integration among elements are found to be nonsignificant in most cases (tables 3, S6). The main exceptions are linked to the weak integration between the femur and humerus in didelphimorphs (table 3) and the strong integration between radius and ulna in diprotodontians (table 3).

Discussion

Tests of Macroevolutionary Hypotheses

The first prediction of the hypothesis of developmentally constrained marsupial forelimbs is that marsupial forelimbs should be less disparate than their hindlimbs. We do not find evidence of this. Disparity values for bones of the fore- and hindlimbs are similar across marsupials and within subgroups (fig. 2). They tend to decrease following a proximodistal gradient with the only exception being the ulna, which is the most disparate limb bone.

The second prediction of the developmental constraint hypothesis states that the strength of integration among bones within the forelimb should be higher than that between fore- and hindlimb. However, our results show that the strength of integration within the forelimb is not different from that between the fore- and hindlimb (table 3; fig. 4). Integration is strongest in general for comparisons among distal elements, both within and between limbs. This pattern resembles that found for other mammals (e.g., Martín-Serra et al. 2015; Hanot et al. 2017).
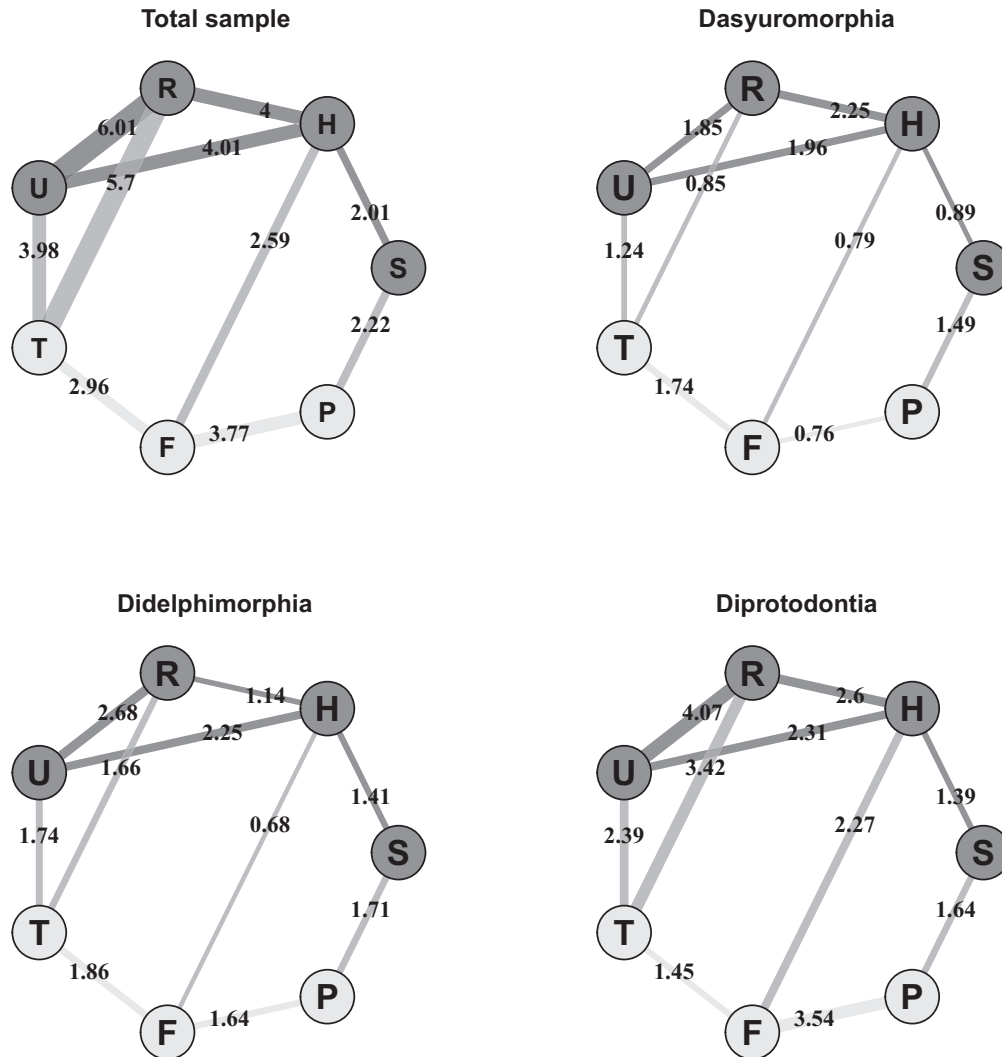
Figure 4: Phylogenetic integration patterns for the total sample of marsupials and for the three orders of marsupials analyzed. Edge thickness and color opacity are proportional to the Z values obtained for each pair of bones (vertices). Z values are also indicated. Dark gray = within forelimb; light gray = within hindlimb; intermediate gray = between limbs. F = femur; H = humerus; P = pelvis; R = radius; S = scapula; T = tibia; U = ulna.

Finally, the third prediction of the constrained forelimb hypothesis is that diprotodontians, which perform the most rigorous crawl as neonates, should show less forelimb disparity and stronger within-forelimb integration than the other groups. Again, our disparity results do not match with this prediction. In fact, diprotodontians have generally higher disparity than the other orders, across both their fore- and hindlimb skeletons and especially in the scapula, humerus, and ulna (table 1; fig. 2). Furthermore, diprotodontians show significantly higher rates of scapular, humeral, and ulnar evolution than didelphimorphs and significantly higher rates of humeral evolution than dasyuromorphs. So our findings regarding disparity do not result from the slightly older clade age of

Diprotodontia. Given that diprotodontians occupy a greater range of body sizes than other marsupials, the greater disparity of diprotodontians could therefore result either from allometry or from the occurrence of more extreme morphologies in large-bodied species compared with small ones (e.g., Stanley 1973). The problem of allometry is addressed by our analyses of the residuals of distance-based regression of shape on size, which also show higher disparity of forelimb element morphologies in diprotodontians compared with other groups (table 1; fig. 2). We also found that small-bodied diprotodontians have disparity values similar to those of small-bodied members of other marsupial groups, with differences being nonsignificant (table 2; fig. 3).

Table 3: Results of the comparison of strength of phylogenetic integration between limb bones for the total sample and the three orders of marsupials analyzed (Dasyuromorphia, Didelphimorphia, and Diprotodontia)

	Z	Pelvis-femur	Femur-tibia	Scapula-pelvis	Humerus-femur	Radius-tibia	Ulna-tibia
Total sample:							
Z	...	3.77	2.96	2.22	2.59	5.70	3.98
Scapula-humerus	2.01	.094	.175	.463	.343	<.001**	.026*
Humerus-radius	4.00	.265	.191	.037	.070	.052	.437
Humerus-ulna	4.01	.353	.259	.056	.102	.026*	.337
Radius-ulna	6.01	.008**	.005**	<.001**	.001**	.490	.073
Dasyuromorphia:							
Z	...	.76	1.74	1.49	.79	.85	1.24
Scapula-humerus	.89	.450	.332	.351	.448	.496	.409
Humerus-radius	2.25	.102	.207	.210	.097	.137	.176
Humerus-ulna	1.96	.171	.326	.325	.166	.220	.277
Radius-ulna	1.85	.173	.320	.319	.168	.219	.274
Didelphimorphia:							
Z	...	1.64	1.86	1.71	.68	1.66	1.74
Scapula-humerus	1.41	.434	.349	.498	.321	.373	.273
Humerus-radius	1.14	.423	.346	.480	.350	.368	.273
Humerus-ulna	2.25	.200	.265	.151	.089	.256	.378
Radius-ulna	2.68	.091	.129	.062	.036*	.127	.218
Diprotodontia:							
Z	...	3.54	1.45	1.64	2.27	3.42	2.39
Scapula-humerus	1.39	.094	.387	.467	.260	.025*	.136
Humerus-radius	2.60	.464	.251	.153	.323	.171	.451
Humerus-ulna	2.31	.342	.337	.230	.432	.111	.351
Radius-ulna	4.07	.153	.053	.016*	.065	.474	.217

Note: Z values for each pair of bones and P values obtained for each comparison between pairs are indicated.

* Statistically significant at a threshold of .05.

** Statistically significant at a threshold of .01.

Comparison of the strengths of limb bone integration among marsupial orders (Dasyuromorphia, Didelphimorphia, and Diprotodontia) shows that within-forelimb integration is not significantly greater than between-limb integration in diprotodontians and that this pattern is similar to that in the other orders (table 3; fig. 4). We therefore do not find evidence that interspecific patterns of diprotodontian forelimb morphology are constrained by their early forelimb development in comparison to other marsupials. In fact, there is no evidence of a differential disparity or strength of integration between fore- and hindlimbs in marsupials or in any of the subgroups analyzed, despite variation in their patterns of early ontogeny (Sears 2004), intraspecific integration (Bennett and Goswami 2011; Kelly and Sears 2011; Garland et al. 2017), and ontogenetic shape change (Sears 2004).

Hypotheses of Previous Studies

The hypothesis that developmental factors constrained the evolution of marsupial forelimbs makes several predictions about comparative patterns, for which we find no supporting evidence. This results in potentially important caveats on the role of developmental constraint in limiting the macroevolution of marsupials. Our findings contrast with those of some previous studies that led several researchers to hypothesize the macroevolutionary importance of this developmental constraint to begin with.

Sears (2004) conducted interspecific comparisons of linear measurements for 66 adult marsupials and 109 adult placental mammals. She found that marsupial scapulae were less disparate than placental scapulae, whereas their pelves were not. Furthermore, she explored ontogenetic patterns of morphological variation in the scapula for four marsupial and 11 placental species, finding that the marsupial species that performed the most extensive crawl to the teat (the diprotodontians *Macropus* and *Trichosurus* and the didelphimorph *Monodelphis*) follow a very similar morphological pattern through ontogeny. In contrast, *Isoodon* (a peramelemorph, in which neonates reach the mother's pouch by slithering a short distance without forelimb action) and most placental mammals show different ontogenetic patterns (Sears 2004). We do not dispute the hypothesis that an early ontogenetic crawl constrains

early ontogenetic patterns of morphological development. However, our results suggest that this ontogenetic constraint has not played a role in structuring macroevolutionary differences in phenotypic disparity among groups of marsupials.

Within-species patterns of integration of the marsupial limb skeleton have received more attention than among-species patterns. Three studies explored patterns of intraspecific integration between limb elements and how they differ between marsupials, placentals, and monotremes (Bennett and Goswami 2011; Kelly and Sears 2011; Garland et al. 2017). These studies analyzed bone lengths (Bennett and Goswami 2011) and two-dimensional landmarks (Kelly and Sears 2011; Garland et al. 2017). Bennett and Goswami (2011) found that within-forelimb integration was greater than within-hindlimb integration for two diprotodontian species (*Macropus giganteus* and *Trichosurus vulpecula*). In contrast, *Didelphis marsupialis* (a didelphimorph) and *Sarcophilus harrisii* (a dasyuromorph) as well as placental mammals exhibited similar strength of integration within both the forelimbs and the hindlimbs. They argued that this difference was due to the more demanding neonate crawl of diprotodontians in comparison to the other two marsupial species (Bennett and Goswami 2011). Kelly and Sears (2011) studied six species of nondiprotodontian marsupials (four didelphimorphs, one microbiotherid, and one paucituberculate), two monotremes, and one placental, finding that integration within the hindlimb was greater than that within the forelimb in four of their six marsupials.

Regarding comparison of within-limb integration to between-limb integration, Bennett and Goswami (2011) found that marsupials showed intraspecific patterns similar to those of placentals. In contrast, Kelly and Sears (2011) found that intraspecific patterns of within-limb integration were generally stronger than between-limb integration in marsupials, although four of their six marsupial species showed significant between-limb covariation for the more distal bones (radius and tibia). Finally, Garland et al. (2017) explored intraspecific integration of four peramelid marsupial species and their sister species *Macrotis lagotis*. As described above, this group shows the least demanding neonate crawl of marsupials for the forelimb, as they barely use them when they slither to reach the mother's pouch. Garland et al. (2017) found greater values of between-limb covariation in peramelemorphs than in previous works on other marsupials (Bennett and Goswami 2011; Kelly and Sears 2011). However, they also suggested that this pattern of integration may be more related to functional adaptation (e.g., mechanical stresses) than with a simple release of the developmental constraints and a return to a pattern determined by serial homology pattern (Garland et al. 2017).

Hypotheses of Developmental Constraints in Marsupials Reconsidered

Taken together, the results of previous studies suggest that intraspecific variation of limb osteology may be structured by early ontogenetic activity in marsupials, with potentially broad implications for mammalian macroevolution. However, patterns at an intraspecific level may not deterministically constrain patterns of macroevolution because functional selective pressures can cause evolution to follow different paths than the lines of least resistance established by the structure of within-species variation (e.g., Renaud et al. 2006). Furthermore, patterns of developmental modularity and integration can also be modified by selection according to the functional interactions of traits (e.g., Wagner and Altenberg 1996; Young and Hallgrímsson 2005; Martín-Serra et al. 2015). In this context, differences between our findings regarding interspecific patterns and previous work on intraspecific patterns suggest that the phenotypic disparity of clades may not be constrained by ontogenetic patterns of variation and furthermore that macroevolutionary (interspecific) patterns of integration are decoupled from those observed within species.

Consequently, although the amount of intraspecific phenotypic variation available for selection in marsupials may be lower for the forelimb than for the hindlimb, it seems that on macroevolutionary scales, selection on the forelimb was nevertheless strong enough to produce similar morphological disparities for both limbs. Indeed, marsupial forelimb bones on average show higher disparity than hindlimb bones (fig. 2). Furthermore, although the fore- and hindlimb may be only weakly integrated with each other during development, we find that they are more strongly integrated during macroevolution. This might result from their coordinated function during quadrupedal locomotion, and the resulting biomechanical requirements may have generated a functionally driven phenotypic integration between limbs (e.g., Garland et al. 2017). Most marsupial species are quadrupedal, so the functional requirements for both limbs are similar, in contrast with other kinds of locomotion (Young and Hallgrímsson 2005). Even large kangaroos, which are widely known bipedal hoppers, move with a pentapedal locomotion (quadrupedal plus the tail) at low speeds, which is their most frequent mode of locomotion (Clancy and Croft 1991).

Our results suggest that the characteristic pattern of development of the marsupial forelimb may not be the primary cause of the reduced phenotypic disparity observed among marsupial species compared with placentals. One alternative explanation is that these causes are related to the evolutionary history of these two groups (Sánchez-Villagra 2013): the more restricted geographical distribution of marsupials compared with placentals during the

Cenozoic may have been the main reasons of the apparent success of placentals in comparison to marsupials. Nevertheless, other explanations are also possible. For example, we are not aware of any systematic comparison of forelimb disparity among marsupial species relative to that of placental species. Assertions of apparently higher disparity of forelimb morphologies in placental mammals generally rely on qualitative observations of extreme morphologies, which result from rare evolutionary transitions to highly distinct ways of life. In any case, this issue warrants continued investigation and may be a pressing target for theoretical advancement in the study of mammalian evolution.

Understanding the tremendous diversity of extant phenotypes is a central goal of comparative biology. A fundamental question in this program asks, What are the drivers and constraints on large-scale phenotypic diversification? Our findings contribute to the classical debate about the interplay between natural selection and developmental constraints (see the introduction). They show that although developmental patterns can have a great influence on phenotypic evolution, other external aspects (such as natural selection, environment, or historical contingency) can play a major role and override the influence of developmental constraints on macroevolutionary time spans.

Acknowledgments

Statement of authorship: A.M.-S and R.B.J.B. designed the research and collected the data, A.M.-S. performed the analyses, and A.M.-S. and R.B.J.B. wrote the article.

Literature Cited

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

———. 2017. Multivariate comparative methods: evaluations, comparisons, and recommendations. Systematic Biology 67:14–31.

Adams, D. C., M. L. Collyer, A. Kaliontzopoulou, and E. Sherratt. 2017. Geomorph: software for geometric morphometric analyses. R package version 3.0.5.

Adams, D. C., and R. Felice. 2014. Assessing phylogenetic morphological integration and trait covariation in morphometric data using evolutionary covariance matrices. PLoS ONE 9:e94335.

Baker, J., A. Meade, M. Pagel, and C. Venditti. 2016. Positive phenotypic selection inferred from phylogenies. Biological Journal of the Linnean Society 118:95–115.

Bennett, C. V., and A. Goswami. 2011. Does developmental strategy drive limb integration in marsupials and monotremes? Mammalian Biology 76:79–83.

Clancy, T. F., and D. B. Croft. 1991. Differences in habitat use and grouping behavior between macropods and eutherian herbivores. Journal of Mammalogy 72:441–449.

Collar, D. C., T. J. Near, and P. C. Wainwright. 2005. Comparative analysis of morphological diversity: does disparity accumulate at the same rate in two lineages of centrarchid fishes? Evolution 59:1783–1794.

Collyer, M. L., D. J. Sekora, and D. C. Adams. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. Heredity 115:357–365.

Dryden, I. L., and K. Mardia. 1998. Statistical analysis of shape. Wiley, Chichester.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

FEI Visualization Sciences Group. 2015. Avizo Lite 9.0.1. FEI, Hillsboro, OR.

Felice, R. N., and A. Goswami. 2018. Developmental origins of mosaic evolution in the avian cranium. Proceedings of the National Academy of Sciences of the USA 115:555–560.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. American Journal of Human Genetics 25:471–492.

Freckleton, R. P. 2012. Fast likelihood calculations for comparative analyses. Methods in Ecology and Evolution 3:940–947.

Futuyma, D. J. 2010. Evolutionary constraint and ecological consequences. Evolution 64:1865–1884.

Galis, F., J. A. Metz, and J. J. van Alphen. 2018. Development and evolutionary constraints in animals. Annual Review of Ecology, Evolution, and Systematics 49:499–522.

Garland, K., A. Marcy, E. Sherratt, and V. Weisbecker. 2017. Out on a limb: bandicoot limb co-variation suggests complex impacts of development and adaptation on marsupial forelimb evolution. Evolution and Development 19:69–84.

Geiger, M., A. M. Forasiepi, D. Koyabu, and M. R. Sánchez-Villagra. 2014. Heterochrony and post-natal growth in mammals—an examination of growth plates in limbs. Journal of Evolutionary Biology 27:98–115.

Gemmell, R. T., C. Veitch, and J. Nelson. 2000. Birth in marsupials. Pages 621–30 *in* Symposium on Physiological and Biological

Adaptations to Australasian and Southern African Environments. Pergamon-Elsevier Science, Cambridge.

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 B 369:20130254.

Goswami, A., V. Weisbecker, and M. R. Sánchez-Villagra. 2009. Developmental modularity and the marsupial-placental dichotomy. Journal of Experimental Zoology 312B:186–195.

Gunz, P., and P. Mitteroecker. 2013. Semilandmarks: a method for quantifying curves and surfaces. Hystrix 24:103–109.

Hallgrímsson, B., K. Willmore, and B. K. Hall. 2002. Canalization, developmental stability, and morphological integration in primate limbs. Yearbook of Physical Anthropology 45:131–158.

Hanot, P., A. Herrel, C. Guintard, and R. Cornette. 2017. Morphological integration in the appendicular skeleton of two domestic taxa: the horse and donkey. Proceedings of the Royal Society B 284:20171241.

Kelly, E. M., and K. E. Sears. 2011. Reduced phenotypic covariation in marsupial limbs and the implications for mammalian evolution. Biological Journal of the Linnean Society London 102:22–36.

Lillegraven, J. A. 1975. Biological considerations of the marsupial-placental dichotomy. Evolution 29:707–722.

Lowe, C. J., and G. A. Wray. 1997. Radical alterations in the roles of homeobox genes during echinoderm evolution. Nature 389:718–721.

Martín-Serra, A., and R. B. J. Benson. 2020. Data from: Developmental constraints do not influence long-term phenotypic evolution of marsupial forelimbs as revealed by interspecific disparity and integration patterns. American Naturalist, Dryad Digital Repository, https://doi.org/10.5061/dryad.900ng75.

Martín-Serra, A., B. Figueirido, J. A. Pérez-Claros, and P. Palmqvist. 2015. Patterns of morphological integration in the appendicular skeleton of mammalian carnivores. Evolution 69:321–340.

Maynard-Smith, J., R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. 1985. Developmental constraints and evolution: a perspective from the Mountain Lake conference on development and evolution. Quarterly Review of Biology 60:265–287.

McKitrick, M. C. 1993. Phylogenetic constraint in evolutionary theory: has it any explanatory power? Annual Review of Ecology, Evolution, and Systematics 24:307–330.

Mitchell, K. J., R. C. Pratt, L. N. Watson, G. C. Gibb, B. Llamas, M. Kasper, J. Edson, et al. 2014. Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. Molecular Biology and Evolution 31:2322–2330.

Nowak, R. M. 1999. Walker's mammals of the world. John Hopkins University Press, London.

O'Meara, B. C., C. Ané, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

R Core Team. 2017. R: a language and environment for statistical computing. Version 3.4.1. R Foundation for Statistical Computing, Vienna. http://cran.R-project.org.

Renaud, S., J. C. Auffray, and J. Michaux. 2006. Conserved phenotypic variation patterns, evolution along lines of least resistance, and departure due to selection in fossil rodents. Evolution 60:1701–1717.

Richardson, M., and A. Chipman. 2003. Developmental constraints in a comparative framework: a test case using variations in phalanx number during amniote evolution. Journal of Experimental Zoology 296B:8–22.

Rohlf, F. J., and M. Corti. 2000. The use of partial least-squares to study covariation in shape. Systematic Biology 49:740–753.

Sánchez-Villagra, M. R. 2013. Why are there fewer marsupials than placentals? on the relevance of geography and physiology to evolutionary patterns of mammalian diversity and disparity. Journal of Mammalian Evolution 20:279–290.

Sears, K. E. 2004. Constraints on the morphological evolution of marsupial shoulder girdles. Evolution 58:2353–2370.

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

Stanley, S. M. 1973. An explanation for Cope's rule. Evolution 27:1–26.

Tyndale-Biscoe, H., and M. B. Renfree. 1987. Reproductive physiology of marsupials. Cambridge University Press, Cambridge.

Veitch, C., C. E. Nelson, and R. T. Gemmell. 2000. Birth in the brushtail possum, *Trichosurus vulpecula* (Marsupialia: Phalangeridae). Australian Journal of Zoology 48:691–700.

Venditti, C., A. Meade, and M. Pagel. 2011. Multiple routes to mammalian diversity. Nature 479:393–396.

Wagner, G. P. 1988. The influence of variation and development constraints on the rate of multivariate phenotypic evolution. Journal of Evolutionary Biology 1:45–66.

Wagner, G. P., and L. Altenberg. 1996. Complex adaptations and the evolution of evolvability. Evolution 50:967–976.

Washburn, S. L. 1946. The sequence of epiphysial union in the opossum. Anatomical Record 95:353–363.

Young, N. M., and B. Hallgrímsson. 2005. Serial homology and the evolution of mammalian limb covariation structure. Evolution 59:2691–2704.

Zelditch, M. L., D. L. Swiderski, H. D. Sheets, and W. L. Fink. 2012. Geometric morphometrics for biologists: a primer. 2nd ed. Elsevier/Academic Press, Amsterdam.

References Cited Only in the Online Enhancements

Geiser, F. 1986. Thermoregulation and torpor in the kultarr, *Antechinomys laniger* (Marsupialia: Dasyuridae). Journal of Comparative Physiology B 156:751–757.

Janson, C. H., and L. H. Emmons. 1990. Ecological structure of the nonflying mammal community at Cocha Cashu biological station, Manu National Park, Peru. Pages 314–338 *in* A. H. Gentry, ed. Four Neotropical rainforests. Yale University Press, New Haven, CT.

Oakwood, M. 2002. Spatial and social organization of a carnivorous marsupial *Dasyurus hallucatus* (Marsupialia: Dasyuridae). Journal of Zoology 257:237–248.

Paradis, E. 2012. Analysis of phylogenetics and evolution with R. 2nd ed. Springer, New York.