PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Weisbecker V, Beck RMD, Guillerme T, Harrington AR, Lange-Hodgson L, Lee MSY, Mardon K, Phillips MJ. 2023 Multiple modes of inference reveal less phylogenetic signal in marsupial basicranial shape compared with the rest of the cranium. *Phil. Trans. R. Soc. B* **378**: 20220085. https://doi.org/10.1098/rstb.2022.0085

Received: 15 August 2022
Accepted: 17 December 2022

One contribution of 13 to a theme issue 'The mammalian skull: development, structure and function'.

Subject Areas:
evolution, palaeontology, taxonomy and systematics

Keywords:
geometric morphometrics, marsupials, phylomorphospace, parsimony, phylogenetics

Author for correspondence:
Vera Weisbecker
e-mail: vera.weisbecker@flinders.edu.au

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6566745.

THE ROYAL SOCIETY PUBLISHING

Multiple modes of inference reveal less phylogenetic signal in marsupial basicranial shape compared with the rest of the cranium

Vera Weisbecker[1], Robin M. D. Beck[2], Thomas Guillerme[3], Arianna R. Harrington[4], Leonie Lange-Hodgson[5], Michael S. Y. Lee[1,7], Karine Mardon[6] and Matthew J. Phillips[8]

[1]College of Science and Engineering, Flinders University, Adelaide, South Australia 5042, Australia
[2]School of Science, Engineering and Environment, University of Salford, Salford, M5 4WT, UK
[3]School of Biosciences, University of Sheffield, Sheffield, S10 2TN, UK
[4]Department of Biology, Southern Utah University, Cedar City, UT, 84720, USA
[5]School of Biological Sciences, and [6]Centre of Advanced Imaging, University of Queensland, Saint Lucia, Queensland, 4072, Australia
[7]Earth Sciences Section, South Australian Museum, Adelaide, South Australia, 5000 Australia
[8]School of Biology & Environmental Science, Queensland University of Technology, Brisbane, Queensland, 4000, Australia

VW, 0000-0003-2370-4046; RMDB, 0000-0002-7050-7072; TG, 0000-0003-4325-1275; MSYL, 0000-0002-3905-0887; MJP, 0000-0003-1532-449X

Incorporating morphological data into modern phylogenies allows integration of fossil evidence, facilitating divergence dating and macroevolutionary inferences. Improvements in the phylogenetic utility of morphological data have been sought via Procrustes-based geometric morphometrics (GMM), but with mixed success and little clarity over what anatomical areas are most suitable. Here, we assess GMM-based phylogenetic reconstructions in a heavily sampled source of discrete characters for mammalian phylogenetics—the basicranium—in 57 species of marsupial mammals, compared with the remainder of the cranium. We show less phylogenetic signal in the basicranium compared with a 'Rest of Cranium' partition, using diverse metrics of phylogenetic signal (K_{mult}, phylogenetically aligned principal components analysis, comparisons of UPGMA/neighbour-joining/parsimony trees and cophenetic distances to a reference phylogeny) for scaled, Procrustes-aligned landmarks and allometry-corrected residuals. Surprisingly, a similar pattern emerged from parsimony-based analyses of discrete cranial characters. The consistent results across methods suggest that easily computed metrics such as K_{mult} can provide good guidance on phylogenetic information in a landmarking configuration. In addition, GMM data may be less informative for intricate but conservative anatomical regions such as the basicranium, while better—but not necessarily novel—phylogenetic information can be expected for broadly characterized shapes such as entire bones.

This article is part of the theme issue 'The mammalian skull: development, structure and function'.

1. Introduction

Much of the phylogeny of modern mammals appears to be robustly resolved, in large part owing to advances in molecular phylogenetics [1,2]. However, there is a pressing need to better integrate morphological characters into this molecular framework, particularly features of the skeleton and dentition. These are the best-studied anatomical systems in mammals, and also by far the most likely to fossilize [3–5]. In particular, the inclusion of fossil taxa, which typically do not preserve DNA and so must be placed in a phylogeny based on morphological

evidence, may be critical for correct inference of topologies at deeper nodes [5], trait evolution (e.g. [6]), biogeography (e.g. [7]), divergence times [1], and diversification rates (e.g. [8]).

Morphological datasets for phylogenetic analysis generally discretize morphological variation into discrete (binary or multi-state) characters, which are scored in representatives of a clade of interest; continuous data (such as linear measurements or ratios) can also be used (reviewed by [9]), but this is much less common. However, despite their wide use for dating and placing fossils, discrete morphological characters can potentially represent oversimplifications of continuous morphological variation (e.g. [10]).

Geometric morphometrics (GMM) is an approach where Cartesian coordinates of generally homologous landmarks are scaled to the same size and then superimposed, typically using Procrustes methods [11]. Phylogenetic signal is generally present and routinely adjusted for in GMM-based data of mammalian skeletal evolution (e.g. [12–14]). GMM is also an established technique to assess taxonomic boundaries between closely related taxa [15–17]. It is therefore unsurprising that GMM has long been discussed as a potential avenue for incorporating new information into phylogenetic analyses [18], even though its use for direct phylogenetic inference has been criticized as methodologically dubious and unreliable [19–22]. However, this criticism has not prevented diverse attempts to integrate GMM data into phylogenies [23–28], even though few studies have been successful at reflecting known phylogenetic relationships well [18,28–30].

As with discrete characters, it is likely that the shape of some morphological regions contains more phylogenetic information than others [31,32]. Exploration of candidate regions represents an important challenge because GMM data are a fundamentally different modality compared with discrete and univariate continuous data [33]. For example, conventional phylogenetic scoring emphasizes only a subset of observed variability, which can be readily encapsulated as discrete character states. It is the product of human decisions of what 'counts' as a character, how to divide it up into states, and how the scoring proceeds [9,34]. GMM also depends on human landmark choices, but reflects the relative spatial position of the landmarks to each other. In addition, the Procrustes superimposition relates all coordinates to each other, distributing the variance from each point to all points. As a result, the main shape variation is more likely to be captured even if some landmarks are missed or if suboptimal landmarking is employed [16,35]. How this property impacts on the phylogenetic information in a dataset is not well understood, and needs to be explored as the field moves towards the inclusion of GMM into phylogenetic investigations.

The basicranial region of mammals (including the ventrolateral braincase and auditory region) represents an excellent challenge for GMM-supported phylogenetic inference because it is one of the most important sources of discrete phylogenetic characters [36–38]. It includes multiple foramina transmitting nerves, arteries and veins [38], houses the middle ear with its intricate soft-tissue environment [39] and is embryologically highly complex [40]. Because of this, it has been argued that the mammalian basicranium is highly conserved [41,42], whereas other characters such as dentition appear evolutionarily more labile [31]. Further, because the function of the auditory region appears to be largely independent of its precise bony composition (for example, in the bones making up the ossified auditory bulla, if present), at least some changes in this region may be selectively neutral (e.g. [42]). Thus, basicranial morphology should be expected to change slowly, and to show lower levels of homoplasy than other anatomical regions that are less complex and/or are under tighter functional constraint. Basicranial characters have thus been widely used in morphological phylogenetic analyses of broad-scale mammalian relationships [39,43], as well as less inclusive mammalian clades (e.g. [36,44,45]).

Its long history as a source of discrete phylogenetic characters makes the basicranium an interesting test case for whether its shape performs well in reflecting phylogenetic information when approximated as GMM coordinates. Reports on this are currently mixed. In a sample of marmots (genus *Marmota*), ventral skull shape (including the basicranium) had higher phylogenetic signal than the mandible and molars, but only when considering recent divergences (within time scales of a few million years or less). A study on ursids (bears) [24] suggested limited amounts of phylogenetic information at deeper nodes close to the first divergence of the ursid clade in the late Palaeogene. Lastly, an earlier study [46] on papionin primates suggested that discrete characters are far superior to GMM-based basicranial shape variation for accurately resolving phylogenetic relationships. By contrast, GMM of the basicranium performed better than other parts of the skull in resolving relationships among cercopithecine primates [18].

In this study, we evaluate the phylogenetic information of geometric morphometrics-defined basicranial shape in extant marsupial mammals as reflected by diverse analytical methods. We compare the signal in the 'Basicranium' partition (defined here as the region of the middle ear and hypotympanic floor, glenoid fossa and foramina of the ventral braincase) with the signal in the 'Rest of Cranium' (snout, dorsal braincase and zygomatic arches). In contrast to the Basicranium partition, the Rest of Cranium partition has been considered to vary strongly with functional selection [47–49], which can be a driver of homoplasy (e.g. [47,50,51]) and thus reduce the phylogenetic signal present. The parts of the skull that make up the Rest of Cranium partition also tend to be underrepresented in discrete morphological datasets for mammals. For example, in their phylogenetic analysis of marsupials, Beck *et al*. [36] scored 95 morphological characters from the entire cranium (i.e. excluding the mandible and dentition), of which 44% (42/95) came from the basicranium (including the ossicles), but only 11% (10/95) from the dorsolateral braincase, even though these two regions are similar in size. Marsupial mammals are a morphologically diverse mammalian clade for which detailed discrete character matrices have been developed over several decades (reviewed in [36,52]). Here, we 'translate' these discrete characters into approximately equivalent landmark configurations. We then test the expectation that basicranial shape should contain more phylogenetic information than the rest of the cranium at both shallow and deep nodes within marsupial phylogeny. We chose methods from two contexts where phylogenetic signal of shape is relevant: comparative evolutionary studies where phylogenetic signal is evaluated as a potential confounding factor in patterns of adaptation, and clustering approaches that are more reflective of potential tree-reconstruction methods that might incorporate GMM data.

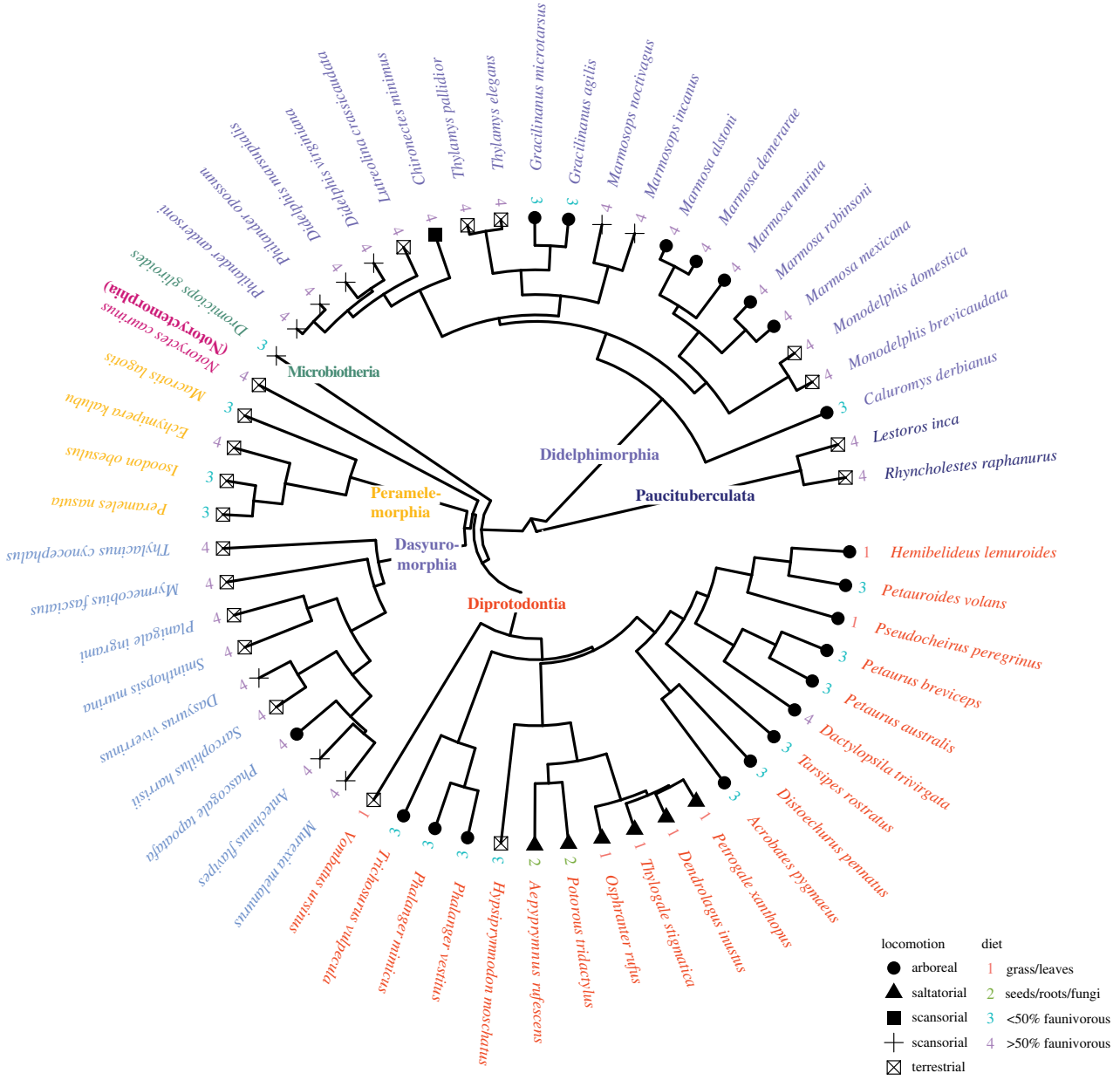

Figure 1. Phylogeny of the species sampled as reconstructed by the molecular 'reference' tree used in this study (see also electronic supplementary material, S1). Branch lengths are proportional to divergence times.(Online version in colour.)

2. Methods

(a) Three-dimensional data acquisitions

The crania of 63 specimens representing 57 marsupial species (electronic supplementary material, S1) were reconstructed from micro-computed tomography (µCT) scans. Scans were mostly acquired by A.R.H. (using a Nikon Metrology XT H 225 ST at the Duke Shared Materials Instrumentation Facility) and K.M./V.W. (using a Siemens Inveon PET-CT scanner at the Centre for Advanced Imaging, University of Queensland). Exceptions are the scan of the thylacine or Tasmanian tiger, *Thylacinus cynocephalus* (AMNH 35244), which was contributed by Douglass Rovinsky (Monash University), from [53], and common wombat, *Vombatus ursinus* (UMZCA 1010), contributed by Philip Cox (University College London). Only adult specimens were used, as determined by full eruption of all molars. Species were all represented as µCT scans with high resolution (28–53 µm, depending on the species). Mesh files were generated by L.L.-H. from the raw scan data using the three-dimensional image processing software Materialise Mimics [54]. The sample represents all seven extant marsupial orders, includes extensive size variation (from 12 g in *Acrobates pygmaeus* to as much as 95 kg for *Osphranter rufus* [55]), and covers species with highly derived crania, such as *Notoryctes* (the marsupial mole), *Tarsipes* (the honey possum) and *Thylacinus* [36] (figure 1). All meshes are available for unrestricted download on MorphoSource project ID: 000448786, with CT scans available after obtaining museum permission.

(b) Molecular reference tree

We constructed a 63-taxon molecular phylogeny to cover most landmarked species, although some species with little DNA represented had to be entered based on the placement of close relatives or previously published works (see electronic supplementary material, S2), as well as several additional species to increase the accuracy of phylogeny reconstruction [56]. The phylogeny was inferred in MrBayes 3.2.7 [57]; divergence times were estimated using MCMCTree [58] within PAML [58]. For details on DNA sequences used, model partitions and divergence dates, see electronic supplementary material, S2.2.

(c) Discrete character trees

To compare the performance of our GMM-based trees with trees based on discrete characters, we also produced phylogenetic

trees based on the discrete morphological character dataset of [36]. We partitioned this dataset into: (1) a dataset comprising basicranial characters, excluding four characters relating to the ossicles, resulting in 38 characters that correspond to characters 47–56 and 67–88 of the original list of [36]; (2) a dataset comprising characters from the remainder of the cranium, resulting in 53 characters that correspond to characters 1–46 and 89–95 of the original list of [36]; (3) a dataset that corresponds to the original, 'full' 180 character dataset of [36], which comprises cranial, mandibular and dental characters. These discrete datasets comprised 144 taxa [36], but only 45 of these were comparable (even when using congenerics) to the 57 in the GMM dataset. Two types of trees were therefore reconstructed from these discrete datasets, using TNT v. 1.5 [78]: one based on the full sample of 144 taxa, with the resulting tree pruned to the 45 taxa that matched the GMM dataset; and another based on just the 45 taxa that were directly shared with the GMM dataset. For further details on characters, taxon replacements, and tree searches, refer to electronic supplementary material, S2.

(d) Data on cranial function

Locomotor and dietary categories were assigned according to several primary sources [55,59–62] and are summarized in figure 1. For details on these, refer to electronic supplementary material, S2.

(e) Landmarking protocol and quality checks

The landmarking protocol covered two partitions: 'Basicranium' and 'Rest of Cranium', which together comprise the 'Full Cranium' configuration. Landmarks were placed in the Checkpoint software [63] by one operator (L.L.-H.). Basicranial landmarks were determined based on discrete characters from Sánchez-Villagra & Wible [64] and Horovitz & Sanchez-Villagra [52] (see supplementary material, S2 for descriptions of the landmarks, and figure 2 for landmark placements). These included landmarks on structures that are not always visible externally (such as the promontorium of the petrosal, and the transverse canal). These were placed using the slicing windows of Checkpoint, and sometimes with reference to the original CT scans. One of the disadvantages of GMM is that it does not allow for presence/absence data; each landmark must be present in all taxa sampled. We therefore modified the placement of landmarks to characterize the anatomical region corresponding to these landmarks even in the absence of a character. For example, in the case of an intramural connection between the transverse canal foramina, landmarks were placed in close proximity at the lateral extremes of the intramural connection where it exists, or far apart but in the same area when it does not. Placing these landmarks consistently was challenging, which is of note because it may represent a limitation on the success with which GMM protocols can be used to encapsulate complex anatomical variation. Details on decisions are noted in electronic supplementary material, S2. We did not landmark characters that were unique for individual species, or polymorphic within species. Rest of Cranium landmarks were similarly placed based on discrete character descriptions in Horovitz & Sanchez-Villagra [52], with the addition of a few landmarks that represented the overall shape of the cranium (e.g. tip of nasals, suture between maxilla and jugal).

Landmarks were imported into R and processed using the geomorph package v. 4.0.4 [65,66]. Landmarks were scaled and superimposed through generalised Procrustes analysis (GPA), as this remains the most widely used approach in GMM, although other methods might also have merit [33]. Subsequently, coordinates and centroid sizes were averaged for species represented by multiple specimens. A second Procrustes fit was then performed to remove the asymmetric component of cranial shape, such that only the symmetric component was retained for further analysis.

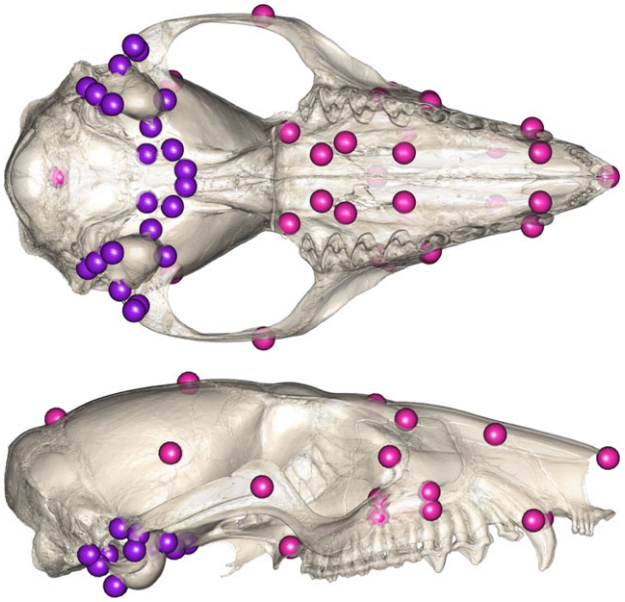

Figure 2. Landmarking protocol plotted onto a surface mesh representing the mean shape of the generalized Procrustes analysis (GPA) of the whole configuration. Purple landmarks represent the Basicranium partition, pink landmarks represent the Rest of Cranium partition.(Online version in colour.)

This was deemed acceptable because a Procrustes ANOVA showed that the effect of 'side' (the asymmetric component) explained less than 0.05% of the variation and was not significant ($p = 0.059$). Alignments were performed separately for the Full Cranium and each partition to ensure that coordinate alignment of each partition was independent from the other. The analyses required for landmark processing, and all subsequent analyses, table outputs and most figures can be replicated by running the code on github repository https://github.com/VWeisbecker/Marsupials_Basicranium_vs_Cranium_Phylosignal.

(f) Analyses of allometry, function, phylogenetic signal and disparity

The shape of the mammalian cranium is often found to be allometric (varying non-uniformly with body mass or cranial dimensions; reviewed in [67]) and thought to be under heavy functional selection [49,51]. This means that phylogenetic signal in shape variation may conceivably be related to variation in size or function, which could lead to homoplasy in shape [31,32,68] and might be more easily evaluated using metrics of size, locomotion, or diet alone. We explored this by asking how much shape variation was associated with overall cranial centroid size (see below) and two widely cited contributors to cranial shape: diet [47,49,51] and locomotion (reviewed in [69]). For this, we used phylogenetically informed generalized least squares (GLS) analysis as implemented in the mvMORPH package [70]. This uses a maximum-likelihood approach to first determine which evolutionary model (Brownian Motion, Ornstein–Uhlenbeck or Early Burst) fitted our models best through comparison of the generalized information criterion (GIC) value from each model fit. Subsequently, we used mvMorph's manova.gls function, which uses a penalized likelihood approach, to ask whether there was a significant association between shape and size in the best-fitting model. The fitting procedure (penalized maximum-likelihood with leave-one-out cross-validation) and MANOVA significance testing (based on Pillai's test statistic) were based on the analyses and instructions in [71] and [72]. We also retained the size-free residuals of shape from the best-fitting model (termed 'residual shape' herein) and ran most following analyses using these

residuals as well as a means of understanding whether the analysis results were dependent on size variation to any substantial degree. All analyses reported in the results were replicated based on these size-adjusted residuals, with similar results.

Variation in size, diet and locomotion in marsupials often arises in unreplicated evolutionary transitions from one state to another [12,55]. This means that phylogenetic correction might underestimate the significance and amount of variation explained by size, locomotion or diet [73]. We therefore also explored non-phylogenetically corrected linear models implemented in geomorph's procD.lm function.

We used geomorph's physignal function for each partition and centroid size of the whole configuration to replicate the commonly used assessment of phylogenetic signal in shape, K_{mult}. This is a multidimensional generalization of Blomberg's K [74], assuming Brownian motion. A K_{mult} value closer to 1 means that shape varies proportionately to the branch lengths of the phylogeny, i.e. it evolves as expected for a Brownian motion process and contains high phylogenetic signal [74]. Lastly, we also measured how much disparity (measured as Procrustes variance, which is a convenient measure of the distance of specimens from the multidimensional mean shape; [75]) was displayed by Basicranium and Rest of Cranium, compared with the Full Cranium. This was the only analysis where partitions were analysed from the Full Cranium GPA alignment.

(g) Phylogenetic aligned principal components analysis

To visually assess the amount of variation that most closely reflects our phylogeny, we computed a phylogenetically aligned principal component analysis, or PACA, in geomorph [76]. This maximizes the co-variation between the first principal component (PC) and the phylogenetic covariance matrix. It also provides partial relative variance (RV) values, and an additional measure of phylogenetic signal as the alignment of each PC with the phylogeny [76]. Lastly, PACA analyses allow an assessment of which independent combinations of landmarks support the phylogenetic differentiation represented by each PAC axis. We also retained the ancestral shape estimations from the PACA output for each node for later use in tree inference.

(h) Phylogenetic and clustering analyses

We created trees of the Full Cranium, Basicranium and Rest of Cranium partitions, as well as the residual shape in each case, using unweighted pair group method with arithmetic mean (UPGMA) cluster analysis and neighbour-joining (NJ) cluster analysis. These allow the rapid inference of trees and have been used in several previous studies that have used GMM data for phylogenetic reconstruction (e.g. [18,25,27]). Cluster trees were created from Euclidean distances of species to each other (equivalent to Procrustes distances) using the phangorn R package [77].

Second, we conducted parsimony-based phylogenetic analysis of our three landmark configurations as implemented in TNT 1.5 [78]. The algorithms used for landmark-based tree searches allow the investigation of individual landmark configurations such as our Basicranium and Rest of Cranium partitions, based on assessment of landmark changes from hypothetical ancestral configurations [79], with the caveat that the investigation of just one partition is probably suboptimal [27]. TNT searches of landmarks weighted all landmarks equally, and involved 50 replicates of tree bisection–reconnection (TBR) on random starting trees, holding up to 300 trees. Owing to the large number of analyses, and the complexity/duration of searches involving three-dimensional character-space, this search strategy was relatively thorough yet feasible (run time less than one week). Trees were rooted to maximize congruence with the reference phylogeny by using *Rhyncholestes raphanurus* as the outgroup taxon (figure 1); however, as the tree-distance metrics used to compare trees compared unrooted networks, outgroup rooting decisions do not affect the results. The executables used for tree inference (TNT datafiles, macros and batch commands and output trees) are given in electronic supplementary material, S3. The TNT analyses considered each shape as one multidimensional configuration, i.e. weighting individual landmarks equally. While multiple configurations are recommended for inferring phylogeny [27], we here focus on how well the Basicranial and Rest of Cranium partitions matched a reference phylogeny. This was to test the expectation that trees based on basicranial shape most closely reflect the topology of the reference phylogeny (similar to previous approaches (e.g. [25,27])). Compatibility between R and TNT inputs/outputs was achieved using Ascarrunz *et al.*'s [23] writeland function and Kranz's ReplaceInFiles Addin (https://github.com/skranz/ReplaceInFiles).

We used the DNA alignments that were used to derive our reference phylogeny to create four 'molecular-based' phylogenetic trees, using both UPGMA and parsimony-criterion tree-building methods. This allowed direct performance comparisons with the three-dimensional landmark data under comparable analytical conditions. UPGMA trees for the DNA data were inferred in PAUP* [80], using three different distance metrics: raw similarity (uncorrected p-distances), likelihood under a general time-reversible substitution matrix (GTR), and likelihood under a GTR matrix with among-site rate variability represented by the gamma and invariant sites parameters (GTRig). The parsimony tree was inferred in TNT v. 1.5, with 100 replicates of TBR on random starting trees, holding up to 1000 trees; rooting was as discussed above for the GMM parsimony trees. Files related to these analyses are in electronic supplementary material, S3.

(i) Topological comparisons of tree distances

We compared the similarity of all trees we generated with the reference molecular phylogeny using generalized Robinson–Foulds (RF) distances implemented in the TreeDist R package [81]. These are based on strict RF metrics (which is a relatively crude metric that counts only the number of splits shared between two trees) moderated by a measure of overall similarity [81]. To contextualize the distances of our reference tree relative to random expectation, we also generated a distribution of distances between 10 000 random trees and the molecular tree (similar to the process described in [25]). If there is no overlap in distances between random versus reference trees compared with GMM-based versus reference trees, the GMM-based tree is thus better than random at 1/10 000 i.e. $p = 0.0001$.

(j) Comparison of GMM-based trees with the discrete cranial phylogeny

To understand how well GMM-based trees perform relative to our discrete cranial phylogeny (see above), we used the same distance-based comparisons as described in the previous paragraph, using reference, UPGMA, NJ and parsimony trees that were pruned to the species represented in the discrete dataset.

(k) Comparison of GMM-based trees with reference phylogenies with increasing node depth

Here, we asked whether the Basicranium and Rest of Cranium partitions differ in their abilities to resolve phylogenetic relationships at deeper nodes compared with shallower ones, based on comparison with our molecular reference tree. We first created datasets of successively pruned trees based on the full reference phylogeny by sequentially collapsing the most recently diverged pair of taxa in the tree (see electronic supplementary material, S4 for a movie of the tree collapse process). Pruning was stopped when the

Table 1. Results from the phylogenetically informed generalized least squares (GLS) and linear model (lm)-based analyses of the association of shape with log-transformed centroid size, locomotion and diet. *F*, *F*-statistic; *p*, significance of association (with a significance threshold of $p < 0.05$); SS, sum of squares.

	Pillai statistic	*p*	SS (lm)	R^2 (lm)	*F* (lm)	*p* (lm)
size						
Full Configuration	0.95	0.000	0.14	0.12	7.4	0.001
Basicranium	0.88	0.000	0.15	0.09	5.31	0.001
Rest of Cranium	0.99	0.001	0.22	0.12	7.15	0.001
locomotion						
Full Configuration	2.74	0.844	0.19	0.15	2.35	0.001
Full Configuration residuals	2.83	0.756	0.17	0.15	2.37	0.001
Basicranium	2.32	0.443	0.24	0.14	2.05	0.001
Basicranium residuals	2.43	0.253	0.23	0.14	2.15	0.002
Rest of Cranium	2.59	0.668	0.33	0.17	2.75	0.001
Rest of Cranium residuals	2.67	0.552	0.29	0.17	2.63	0.001
diet						
Full Configuration	2.2	0.511	0.12	0.1	5.9	0.001
Full Configuration residuals	2.2	0.583	0.11	0.1	6.36	0.001
Basicranium	1.89	0.174	0.16	0.09	5.64	0.001
Basicranium residuals	1.82	0.284	0.16	0.1	6.25	0.001
Rest of Cranium	2.26	0.043	0.22	0.12	7.24	0.001
Rest of Cranium residuals	2.25	0.069	0.16	0.09	5.56	0.001

number of tree tips was less than six. This is because, at these numbers, there are limited combinations of topologies, so that random trees are expected to include all possible trees. We then created matching GMM datasets for both partitions and the Full Cranium (including for size-corrected residuals) for each pruned tree. This was done by replacing the collapsed taxon pairs with the corresponding ancestral estimation for this node (derived from the PACA analysis output as described above). This node estimation can be seen as a weighted average of the states of the descendants of this node. Hence, we are comparing the performance of the two partitions across successively deeper and smaller subsets of nodes. The caveat here is that some 'taxa' in the deeper trees are estimated ancestral states, with decreasing precision at deeper nodes, rather than direct observations. However, even if ancestral state accuracy is a potential confounding factor, it is still valid to compare results between the Basicranium and Rest of Cranium, since they are compared on same basis. Each of these reduced landmark datasets was subjected to the same workflow as above: after cluster and parsimony analysis, distances to the pruned reference tree were compared with a distribution of 10 000 distances of random trees with the appropriate number of tips. To summarize this information visually, at every tree reduction step we assessed whether the distance between the landmark partition-generated trees overlapped with the lowest distance between the reference tree and 10 000 randomly generated distances.

(l) Distance matrix correlations

We used an additional assessment of cophenetic matrix correlation to determine the degree of similarity between all cluster trees (UPGMA, NJ and parsimony) and the trees based on molecular data. Unlike the tree distance metrics, which only assess the correctness of the tree's topology, this allows an assessment of how similar the distances of taxa relative to each other are compared with the reference phylogeny. For this, we created the cophenetic distance matrices for all trees, which reflect the patristic distance of species in terms of branch lengths that separate them. Similar to the comparison with random trees in our tree distance analyses, we used a Mantel test from the vegan package [82] with 10 000 permutations to determine the matrix correlation and its significance between these trees and the reference phylogeny. We also used this approach to compare the degree to which the cluster trees represent the Euclidean distances between species in the coordinate datasets (equivalent to Procrustes distances), as a means of assessing how well the cluster diagrams reflect the morphological data they are based on.

3. Results

(a) Analyses of allometry, function and phylogenetic signal

For all phylogenetically informed generalized least squares models, Ornstein–Uhlenbeck evolution (OU) was the preferred (lowest-GIC) model (electronic supplementary material, S5), and all MANOVAs were therefore computed assuming OU evolution. All analyses showed that size is significantly associated with shape, but in the linear model it accounts for relatively little of the shape variation (a maximum of 12%; table 1). Associations of diet or locomotor mode are significant and explain up to 17% of shape variation in a phylogenetically uncorrected context. The association with shape and diet is significant for the phylogenetically corrected Rest of Cranium analysis, although the linear model R^2 values suggest that diet does not explain much shape variation. GLS analyses of size-adjusted residual shapes yield similar results compared with the full data, confirming that shape variation relative to either of our predictors is not driven by size variation.

Table 2. K_{mult} values for each all landmark configurations, and significance (p) of phylogenetic signal.

	K_{mult}	p
Full Configuration	0.560	0.001
Full Configuration residuals	0.553	0.001
Basicranium	0.398	0.001
Basicranium residuals	0.393	0.001
Rest of Cranium	0.585	0.001
Rest of Cranium residuals	0.570	0.001
log centroid size	0.725	0.001

Both partitions and the Full Configuration contain phylogenetic signal, with little change of K_{mult} when allometric residuals are analysed (table 2). However, the basicranial partition contains less signal than either the Full Configuration or the Rest of Cranium (table 2). The centroid size of the full landmark configuration contains substantially more phylogenetic signal than any of our landmark partitions. The disparity, measured as Procrustes variance, for the full landmark configuration was 0.021. Within the GPA-aligned Full Configuration, the Basicranium partition displayed far less disparity (0.006) compared with the Rest of Cranium (0.016).

The PACA plots (figure 3) for the Full Configuration and Rest of Cranium partitions show much better separations between most marsupial orders than the Basicranium partition. This is also reflected in the relative variance (RV) coefficients for the first and second principal components, which mirror the results for phylogenetic signal (K_{mult}) by being higher in the Full Cranium and Rest of Cranium partition, and lower for the Basicranium. Many ecomorphologically similar clades are placed in disparate parts of the morphospaces in the Full Cranium configuration and Rest of Cranium partition. Examples are peramelemorphians (bandicoots) being well separated from the diprotodontian 'rat' kangaroos, even though species of both clades are scratch-digging and omnivorous and/or fungivorous, and didelphids being well separated from dasyuromorphians, even though species of both clades span a similar size range and are typically faunivorous. Size-residual-based analyses showed largely the same pattern (electronic supplementary material, S6).

(b) Assessing the relative performance of partitions in phylogenetic and clustering analyses

UPGMA, NJ, and parsimony-based trees derived from the Basicranial partition are least similar to the reference phylogeny (i.e. have the highest generalized RF values), with the parsimony-based tree for Basicranium being least similar (figure 4). UPGMA and NJ cluster distances are overall similar, but the Rest of Cranium UPGMA tree was far more similar to the reference topology than all others. DNA data vastly outperform the morphological data, and the parsimony-based DNA tree is topologically identical to the reference (Bayesian) DNA tree; this is expected given they are generated using the same molecular dataset. However, for the DNA data, UPGMA performs far worse than parsimony (likely owing to the large amount of missing data in the dataset [83]). The same result was found in analyses of size-adjusted residuals (electronic supplementary material S6).

(c) Comparison of GMM-based trees with the discrete cranial phylogeny

The parsimony-based 50% majority-consensus tree of discrete data outperformed most GMM-based analyses. However, as with the GMM-based trees, discrete-based trees derived from the basicranial dataset had substantially lower similarity to the reference tree, whereas the dataset of all characters (the full matrix from [36]) resulted in a markedly better-performing tree.

(d) Visual examination of GMM-based UPGMA/parsimony trees

Visualizations of the UPGMA and parsimony trees are consistent with the PACA plots and K_{mult} results. The UPGMA dendrogram of the Rest of Cranium partition differentiates larger, well-established clades particularly well (figure 5; for plots of parsimony-based trees, see electronic supplementary material, S7), retrieving monophyletic Peramelemorphia, Didelphimorphia, Paucituberculata, Petauridae, Macropodoidea and most Dasyuridae. By contrast, the Basicranium partition fails to reflect most clades except for Paucituberculata. However, the UPGMA topologies between and within orders are variable in their retrieval of known phylogenetic relationships. For example, in the Rest of Cranium tree, only some phalangerid and larger petauroid possums cluster together and are placed as the sister group of a monophyletic Macropodoidea. In some cases, close relatives cluster together in the UPGMA analysis (e.g. the two included species of *Didelphis*, and the two included species of *Petaurus*), but many species pairs comprise very distantly related taxa, particularly at the base of the tree (e.g. the grouping of the marsupial mole, *Notoryctes caurinus*, with the numbat, *Myrmecobius fasciatus*).

(e) Comparison of GMM-based trees with reference phylogenies with increasing node depth

The poorer phylogenetic reconstruction from basicranial shape relative to the Rest of Cranium persists at deeper nodes of the tree (figure 6; electronic supplementary material, S6 for size residual-based results), with the estimated accuracy of parsimony-based trees again worse than the cluster trees. The RF distances of landmark-based trees become less distinguishable from random trees as the phylogenies are pruned. This is expected, as fewer tree topologies are available with fewer taxa, and random tree generation is likely to cover more of the possible tree space.

(f) Distance matrix correlations

As with all previous analyses, all trees based on basicranial partitions have substantially lower matrix correlations with the reference phylogeny (table 3). As with the tree distance comparisons, UPGMA clusters of the Rest of Cranium partition perform best, followed by the Rest of Cranium or Whole of Cranium NJ and parsimony clusters. Interestingly, the NJ clustering reflect the Procrustes distances (i.e. the morphological distances in their original form) better than

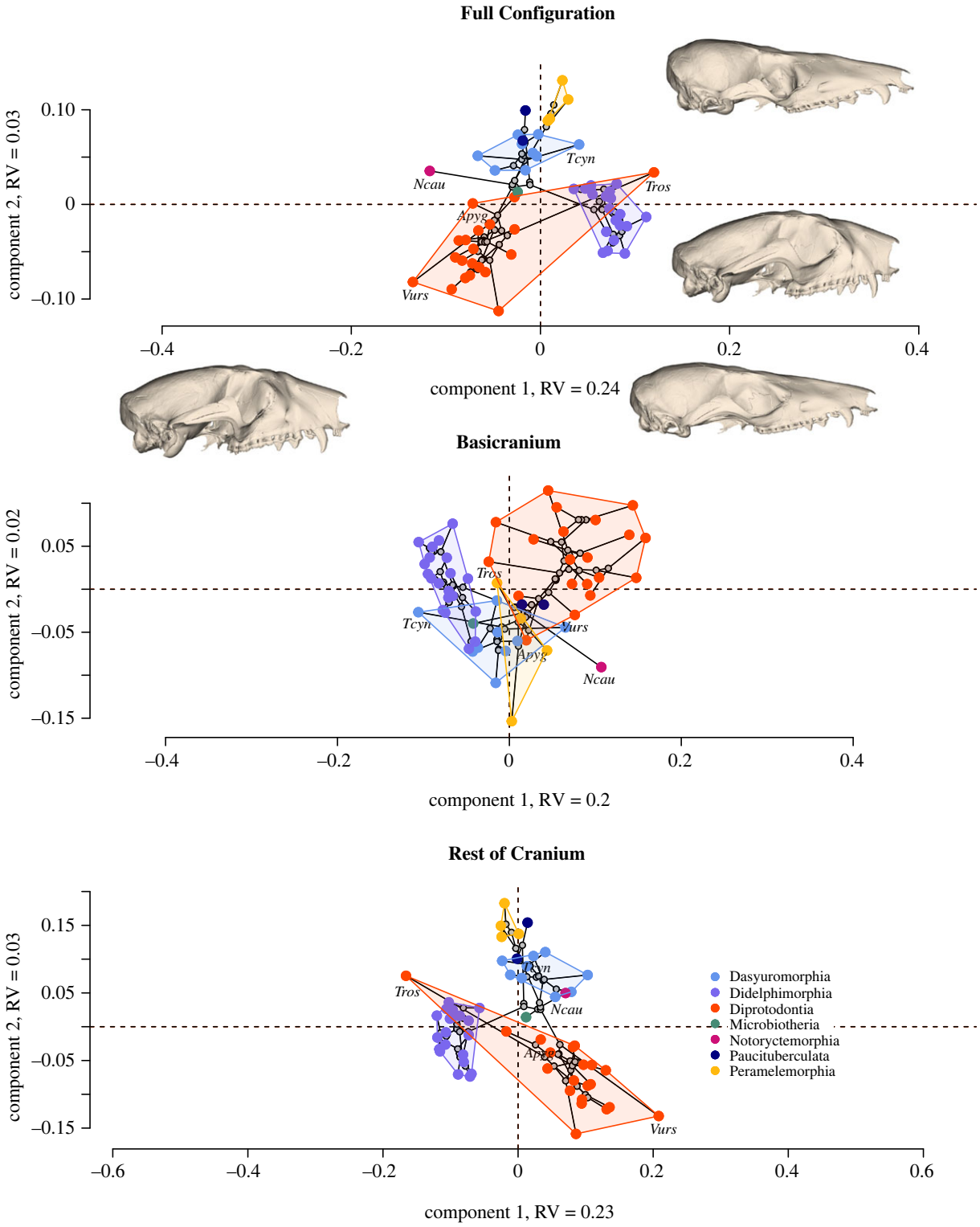

Figure 3. Phylogenetically aligned PCA plots for the Full Configuration and both partitions, with visual representations of the hypothetical whole cranial shapes at PAC axis extremes. RV, relative variance. (Online version in colour.)

UPGMA clustering, despite having a substantially lower matrix correlation with the reference phylogeny (table 3). Parsimony-based analyses reflect the morphological data least well. As with the tree distance comparisons, the parsimony-based analysis of molecular data reveal a highly similar tree to the reference phylogeny, with a matrix correlation of 0.96. However, the UPGMA-based molecular trees perform no better than the best of the Rest of Skull or Full Configuration cluster trees (table 3). Thus, UPGMA trees of molecular data perform better at retrieving topologies than branch lengths.

4. Discussion

Our results show that geometric morphometric landmarks on the marsupial basicranium—a focal area for conventional phylogenetic scoring using discrete characters—provide substantially less phylogenetic resolution than the Rest of Cranium partition, which is typically accorded much less attention in discrete character scoring. Our results add to previous findings of limited phylogenetic signal in the basicranium in analyses of GMM configurations [15,24,46]. Unexpectedly, however, the Basicranium partition also performed much

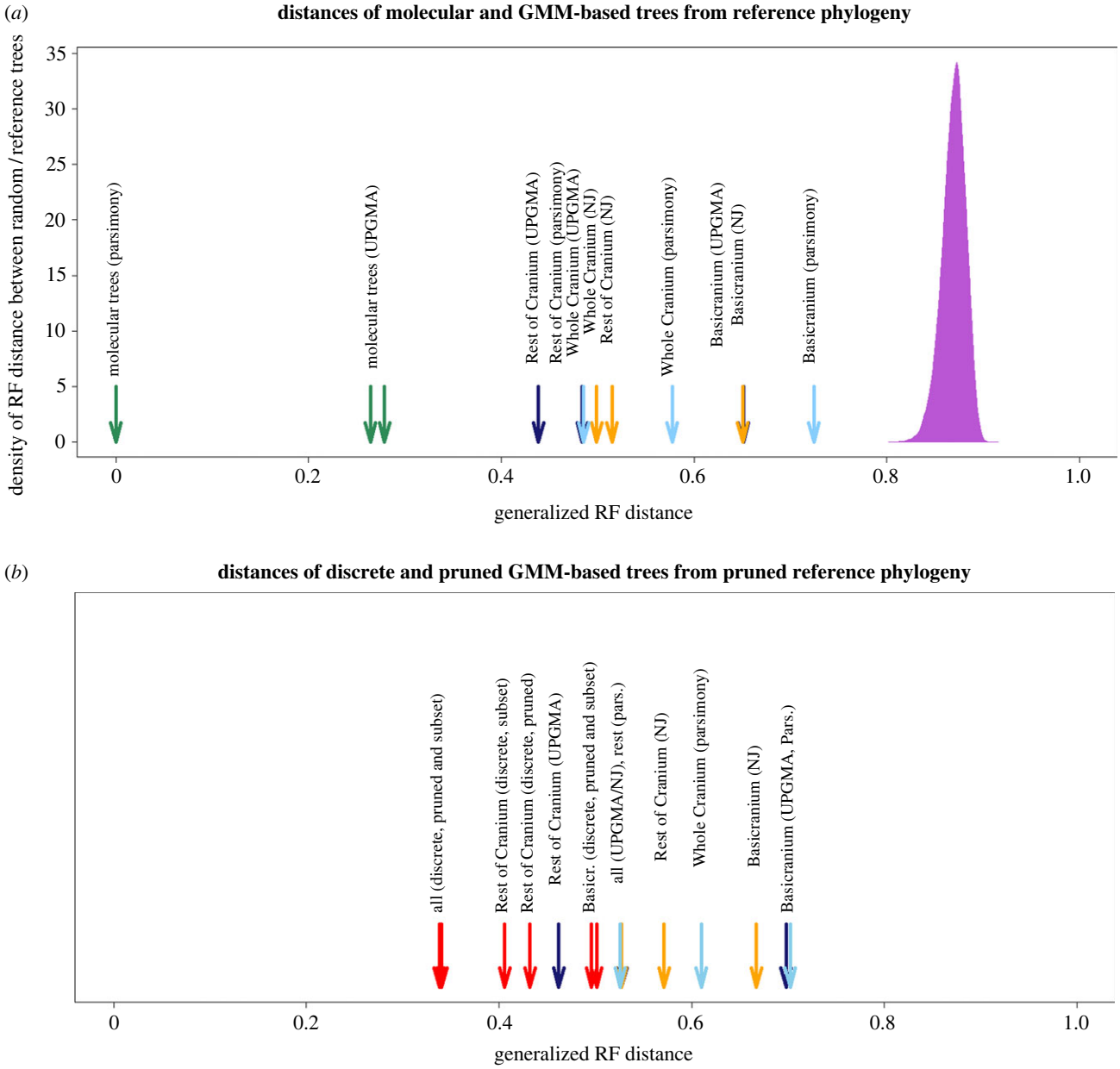

Figure 4. (*a*) Generalized Robinson–Foulds (RF) distances of NJ (orange), UPGMA (dark blue), parsimony-based (light blue) and molecular-based (green) trees from the reference molecular phylogeny, compared with the density distribution of distances of a randomly generated sample of 10 000 random trees (purple curve). (*b*) Generalized RF distances of trees based on discrete data (red), compared with the above trees reduced to the same sample. 'pruned', discrete-based parsimony trees derived from a larger sample and pruned to the dataset matching the GMM data; 'subset', discrete-based parsimony trees derived from a dataset matching the species in the GMM dataset. (Online version in colour.)

worse in our discrete character-based tree comparisons, although these performed better than most GMM-based trees overall.

It is possible that evolutionary conservatism of the basicranium—the very feature that makes it a popular choice for discrete character scoring—in fact reduces the phylogenetic information reflected in this region. In a GMM context, a conserved or constrained morphology would result in less available shape variation (as seen in the lower disparity of the Basicranium compared with the Rest of Cranium partition) and less freedom for landmarks to vary relative to each other. The lower phylogenetic signal in the basicranium is also consistent with the expectation that it has highly integrated constituent parts that are not free to vary relative to each other through evolutionary time. High integration has been shown to substantially reduce phylogenetic signal in simulations [19], and this effect might be particularly strong in the relatively conservative [84] basicranium of marsupials and other metatherians.

The intricate topology of the basicranium also highlights the different properties of GMM versus discrete data acquisitions, as outlined in the Introduction. Discretization of morphological characters draws unambiguous distinctions between one morphology and another, and is easier when variability is relatively straightforward to encode (e.g. presence/absence, or single dimension). This is certainly the case for the basicranium: for example, in terms of discrete basicranial characters, diprotodontians are different from peramelemorphians, and dasyurids are highly distinct compared with didelphids [5]. By contrast, peroryctid peramelemorphians, most didelphids, and *Thylacinus* share a somewhat similar auditory region in terms of the relative shapes and sizes of the tympanic processes of the alisphenoid and petrosal [5] and might therefore be expected to cluster together. Thus,

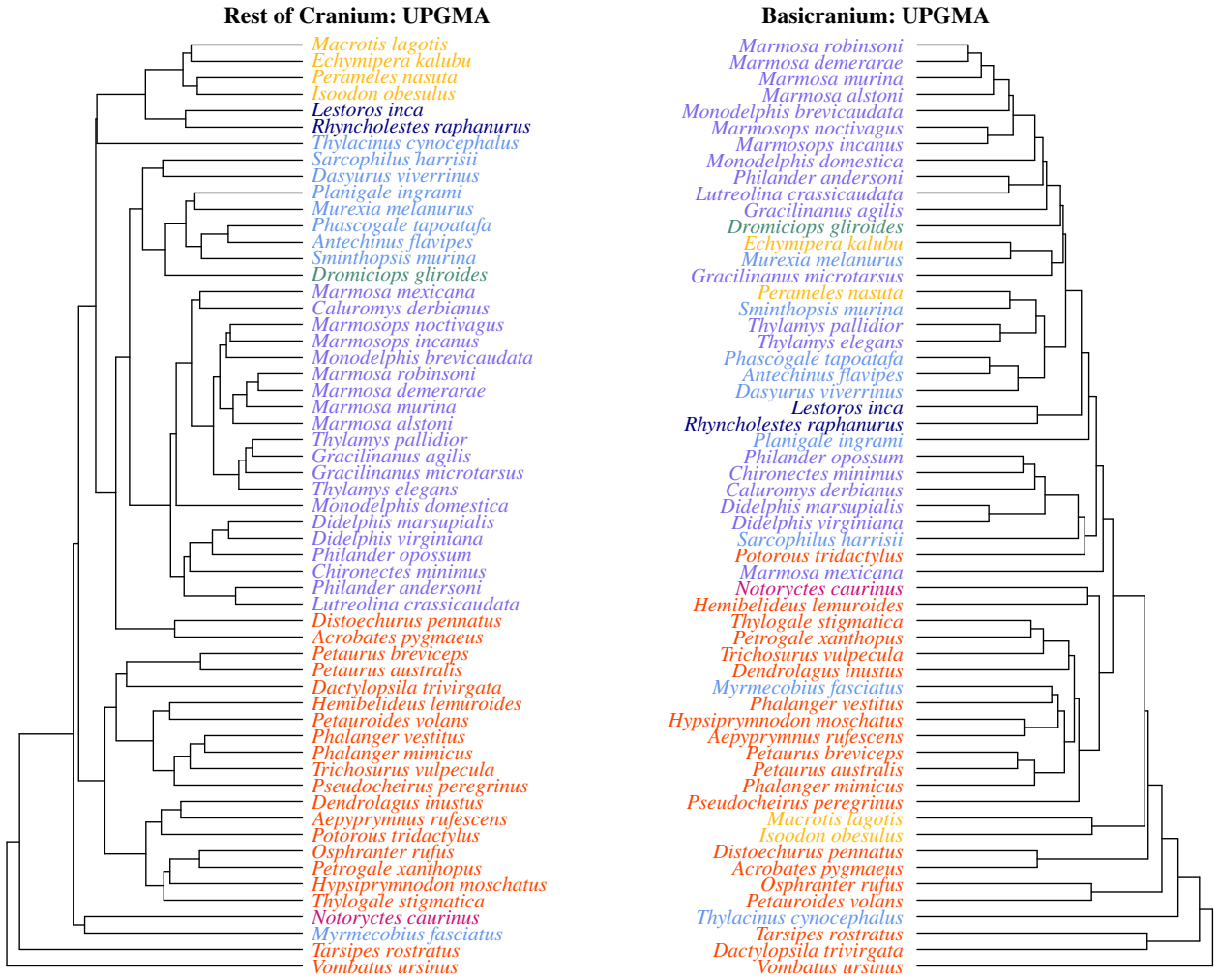

Figure 5. UPGMA-based dendrograms (trees) for all species sampled. Colours indicate orders (figure 3). UPGMA clustering is presented because the UPGMA Rest of Cranium cluster was closest to the reference phylogeny.(Online version in colour.)

Table 3. Matrix correlations determined by Mantel tests of distance matrices as outlined in the top row of the table heads. statistic, Mantel statistic; GPA, based on generalized Procrustes alignment; resids, based on allometrically corrected residuals of shape; pars., parsimony; NJ, neighbour-joining; GTR, general time-reversible substitution matrix of molecular data; GTRig, GTR matrix with among-site variability represented by the gamma and invariant sites parameters. For completeness, distance matrix comparison of the molecular-based parsimony/UPGMA trees is also displayed at the bottom of the second-last column to the right. All correlations were significant at $p < 0.001$.

	patristic vs Procrustes distances		patristic vs mol. reference		Procrustes/mol. distances vs mol. reference		
	statistic: GPA	statistic: resids	statistic: GPA	statistic: resids		statistic: GPA	statistic: resids
pars., Full Config.	0.81	0.83	0.45	0.58	Full Config.	0.46	0.5
pars., Basicr.	0.72	0.72	0.29	0.28	Basicr.	0.29	0.29
parsimony, Rest of Cranium	0.77	0.79	0.59	0.59	Rest of Cranium	0.5	0.51
NJ, Full Config.	0.91	0.94	0.47	0.46			
NJ,Basicr.	0.91	0.92	0.22	0.2			
NJ,Rest of Cranium	0.92	0.91	0.53	0.52	mol., pars.	0.96	
UPGMA, Full Config.	0.83	0.86	0.57	0.59	mol., raw dist.	0.56	
UPGMA, Basicr.	0.83	0.86	0.25	0.29	mol., GTR	0.56	
UPGMA, Rest of Cranium	0.83	0.8	0.63	0.64	mol., GTRig	0.57	

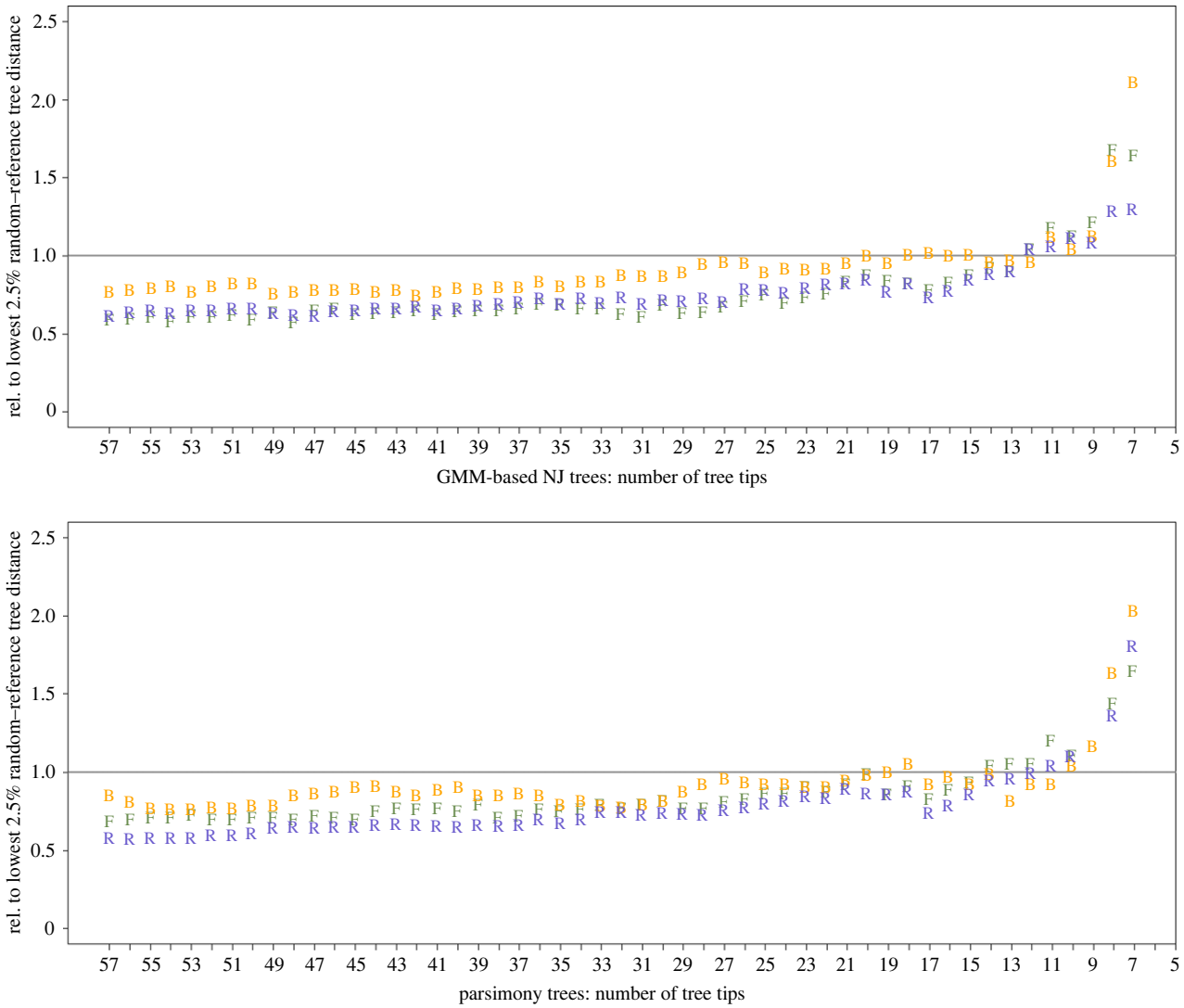

Figure 6. Distances of the NJ and parsimony-based collapsed trees for the three partitions from the reference tree, in relation to whether the distance overlaps with the smallest distance among the distribution of distances between random trees and the true tree. A value of 1 or higher means that a morphological tree overlaps with random trees that have the least distance from the true phylogeny; the smaller the value, the closer the UPGMA/TNT trees are to the molecular tree (0 being 100% congruence/zero distance). F, Full Cranium; B, Basicranium; R, Rest of Cranium. (Online version in colour.)

discrete characters have few degrees of freedom and are somewhat arbitrary yet often infinitely extensible to new taxa and features. By contrast, GMM protocols have less ability to incorporate this discrete information because GMM is limited to common homologous points across the sample. For example, while we could landmark the extent of the tympanic process of the alisphenoid because it contributes to the hypotympanic floor in most species of our sample, we could not account for the diverse composition of the hypotympanic floor among diprotodontians [5]. Presence/absence characters need to be ignored or heavily modified (see Methods), and other character states, such as the extent of the rostral and caudal tympanic process in dasyurids (in which they are fused into a continuous 'petrosal plate' [5]), were simply difficult to observe and are likely prone to error. The reliability of the landmark protocol can of course be improved by using fewer and more easily placed landmarks, but these will inevitably ignore the variation that has led to the widespread use of the basicranium in discrete phylogenetic character scoring. Overall, therefore, the emphasis of conventional discrete character coding on intricate but conservative anatomical details may not be fully translatable into equivalent GMM-based characters (see also [33]).

Cluster reconstructions of the Rest of Cranium partition were particularly successful at capturing the evolutionary transitions between marsupial orders, and this partition also performed much better in our discrete character-based trees. Clustering of ecomorphologically distinctive major clades is common in marsupial GMM datasets [12,13,85], in other mammals (e.g. [49,86–88]), and even in major amniote clades [89]. True to the spreading of variation during Procrustes superimposition, GMM thus seems to capture variation in *Gestalt* (the invariants, or commonalities, of a collection of patterns; [90]) which often accompanies ecomorphological divergences such as those among marsupial orders [91]. It is possible that the stronger signal of the Rest of Cranium partition arises from the evolution of its multiple functional and/or developmental modules [92]. These might be more likely to combine into a unique and clade-specific shape than the spatially more restricted basicranium, regardless of whether the data are GMM-based or discrete. Importantly, the phylogenetic signal of the Rest of Cranium partition was not confounded by common determinants of cranial shape, such as allometry, locomotion, and diet. Our PACA visualizations support this by clearly distinguishing between clades that share similar diets (such as the largely faunivorous dasyuromorphians and

didelphimorphians), an effect also found recently based on linear measurements of marsupial crania [13]. This adds to existing evidence that convergence may be less common than might be expected in cranio-mandibular shape variation [93–95]. It is also potentially a lesser issue in multidimensional contexts because of the many ways that shape can diverge outside of the variation explained by convergence [19]. The capture of overall shape, which is difficult to achieve in a discrete character context, is therefore a potential advantage of GMM-supported phylogenetic analysis.

Unfortunately, the Rest of Cranium captured the least controversial phylogenetic groupings, with obvious, albeit not easily discretized, morphological differences. It would thus not be expected to add further resolution to taxa that have been persistently difficult to place. However, it is possible that a more extensive characterization of this partition—for example through curve or surface semilandmarks—might improve on phylogenetic resolution. Higher dimensionality has been found to improve phylogenetic signal [19], although it is possible that this effect might arise from redundant information in the case of high-density landmark coverage. It should also be added that GMM-based phylogenetic resolution among species that are closely related (such as congenerics, like the multiple representatives of *Marmosa* in our sample) may not be possible. This is because large sample sizes are required to distinguish close relatives [96], but only tree tips, and therefore samples of $n = 1$, can be easily analysed in a phylogenetically informed GMM context.

Our centroid sizes contained far higher phylogenetic signal than the shape variation (similar to results in [12,48]), but did not explain much of the shape variation. Centroid sizes broadly relate to the body mass of a species, which contains strong phylogenetic signal across mammals (e.g. [97]). Separating the phylogenetic signal of shape and size through generalized Procrustes analysis therefore may provide an additional source of data that can be reasonably independent from shape variation, and could be an additional reason to use GMM data in phylogenetic analysis.

The use of GMM data in phylogenetic reconstructions has been quite controversial [19–22,98]. However, the robust retrieval of similar phylogenetic signal among our methods lends encouraging empirical support, in line with some previous work [18,27,29]. The methods had slightly different properties in our analyses, which might make some more suitable than others for estimating phylogenetic signal in GMM data [19]; for example, UPGMA retrieved phylogenetic relationships better than NJ or parsimony, but NJ reflected Procrustes distances between species better. However, assessments of phylogenetic signal in GMM datasets are likely overall robust to the diverse alternatives in nearly every step of the analytical workflow (e.g. [15,18,23–25,27]), e.g. in regards to clustering algorithms, tree comparison metrics and tree-distance- versus distance-matrix-based comparisons. An exception might be squared-change parsimony, which performed worst at retrieving topologies, but reflected patristic distances as well as UPGMA and NJ clusters (but see simulations in [19]). Furthermore, parsimony has some potential advantages, such as being able to co-analyse the GMM data with discrete and/or molecular data.

The agreement of results also suggests that the easily computed K_{mult} metric or PACA visualizations from the geomorph R package could be a convenient and reasonable (if possibly not highly accurate) means of assessing phylogenetic signal in candidate landmark protocols, and where it might be located. These are currently based on the potentially over-simplistic assumption of a Brownian evolution [71], but clearly represent a good approximation of phylogenetic 'usefulness' in our study and others. For example, in contrast to our results and those of others [15,24,46], Cardini & Elton [18] reported high K_{mult} in the basicranial region of cercopithecines compared with other parts of the cranium, and this partition also resulted in better phylogenetic trees. The K_{mult} statistic of mammalian mandible shape also varies widely depending on the clade investigated [48,99–101].

5. Conclusion

Contrary to the focus on anatomical detail in conserved regions in conventional phylogenetic analyses, our results suggest that GMM-encoded shapes of whole bones of relatively variable regions—such as the Rest of Cranium partition—might be more phylogenetically informative than small regions, even at earlier evolutionary divergences. In addition, the low phylogenetic information content in the Basicranium partition confirms that a vetting process for candidate regions is important to avoid the laborious acquisition of relatively uninformative areas, if the primary aim is phylogenetic reconstruction [31].

Our finds are consistent with statistically significant phylogenetic signal reported in whole bones across mammalian crania, mandibulae and postcrania (e.g. [86,87,102]), confirming that skeletal shape is a potential source of phylogenetic information. However, the potential for integrating GMM data into phylogenies is limited by the availability of meaningful avenues for inclusion into phylogenies, particularly because the main technique—parsimony-based analysis—performed worse than our simpler clustering approaches at retrieving topologies. In addition, morphological information is particularly relied upon in the placement of fossils, but undistorted and intact whole bones are uncommon in the fossil record. Complete GMM information for individual bones is therefore rarely achievable in analyses of multiple fossils. Thus, GMM will mostly be an option in specific, uncommon cases where 'overall bone shape' is available and can be contextualized with data from other sources or constraints on the phylogeny (such as successfully done by Parins-Fukuchi [29]).

Data accessibility. The three-dimensional meshes and checkpoint files used to gather landmarks are available under doi:10.25451/flinders.20459793. All other files required to perform the analyses are available on github under https://github.com/VWeisbecker/Marsupials_Basicranium_vs_Cranium_Phylosignal/.

The data are provided in the electronic supplementary material [103].

Authors' contributions. V.W.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, writing—review and editing; R.M.D.B.: investigation, methodology, supervision, validation, writing—review and editing; T.G.: conceptualization, formal analysis, investigation, supervision, validation; A.H.: data curation, funding acquisition, investigation, project administration, writing—review and editing; L.L.-H: data curation, investigation, methodology; M.S.Y.L.: conceptualization, investigation, methodology, writing—review and editing; K.M.: data curation, methodology, resources; M.P.: conceptualization, methodology, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed herein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. V.W. was funded by Australian Research Council Future Fellowship FT180100634. T.G. acknowledges funding by the Royal Society University Research Fellowship UF120016 awarded to Gavin H. Thomas. Funding for CT scans collected by A.R.H. were provided by NSF BCS 1552848, NSF DBI 1458192 and NSF BCS 1825129.

Acknowledgements. Computation was facilitated by a Flinders University Infrastructure grant to M.S.Y.L. Thanks to Douglass Rovinsky for support with figure 1, and to Douglass Rovinksy and Phillip Cox for providing meshes. Access to specimens scanned by A.R.H. were facilitated by Chris Conroy (Museum of Vertebrate Zoology) and Verity Mathis (Florida Museum). The authors acknowledge the facilities and scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at the Centre for Advanced Imaging, University of Queensland. Thanks to reviewer David Polly and an anonymous reviewer.

References

1. Asher R. 2018 Diversity and relationships within crown Mammalia. In *Mammalian evolution, diversity, and systematics* (eds FE Zachos, RJ Asher), pp. 301–352. Berlin, Germany: De Gruyter.
2. Eldridge MDB, Beck RMD, Croft DA, Travouillon KJ, Fox BJ. 2019 An emerging consensus in the evolution, phylogeny, and systematics of marsupials and their fossil relatives (Metatheria). *J. Mammal.* **100**, 802–837. (doi:10.1093/jmammal/gyz018)
3. Guillerme T, Cooper N. 2016 Effects of missing data on topological inference using a total evidence approach. *Mol. Phylogen. Evol.* **94**, 146–158. (doi:10.1016/j.ympev.2015.08.023)
4. Guillerme T, Cooper N. 2016 Assessment of available anatomical characters for linking living mammals to fossil taxa in phylogenetic analyses. *Biol. Lett.* **12**, 20151003. (doi:10.1098/rsbl.2015.1003)
5. Beck RMD, Baillie C. 2018 Improvements in the fossil record may largely resolve current conflicts between morphological and molecular estimates of mammal phylogeny. *Proc. R. Soc. B* **285**, 20181632. (doi:10.1098/rspb.2018.1632)
6. Slater GJ, Harmon LJ, Alfaro ME. 2012 Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* **66**, 3931–3944. (doi:10.1111/j.1558-5646.2012.01723.x)
7. Wisniewski AL, Lloyd GT, Slater GJ. 2022 Extant species fail to estimate ancestral geographical ranges at older nodes in primate phylogeny. *Proc. R. Soc. B* **289**, 20212535. (doi:10.1098/rspb.2021.2535)
8. Rabosky DL. 2010 Extinction rates should not be estimated from molecular phylogenies. *Evolution* **64**, 1816–1824. (doi:10.1111/j.1558-5646.2009.00926.x)
9. Lee MSY. 2020 Clock models for evolution of discrete phenotypic characters. In *The molecular evolutionary clock: theory and practice* (ed. SYW Ho), pp. 101–113. Cham, Switzerland: Springer International Publishing.
10. Wiens JJ. 2001 Character analysis in morphological phylogenetics: problems and solutions. *Syst. Biol.* **50**, 689–699. (doi:10.1080/106351501753328811)
11. Zelditch ML, Swiderski DL, Sheets HD. 2012 Introduction. In *Geometric morphometrics for biologists*, 2nd edn (eds ML Zelditch, DL Swiderski, HD Sheets), pp. 1–20. San Diego, CA: Academic Press.
12. Weisbecker V *et al.* 2021 Global elongation and high shape flexibility as an evolutionary hypothesis of accommodating mammalian brains into skulls. *Evolution* **75**, 625–640. (doi:10.1111/evo.14163)
13. Giannini NP, Morales MM, Wilson LA, Velazco PM, Abdala F, Flores DA. 2021 The cranial morphospace of extant marsupials. *J. Mamm. Evol.* **28**, 1145–1160. (doi:10.1007/s10914-021-09589-y)
14. Goswami A. 2007 Phylogeny, diet, and cranial integration in australodelphian marsupials. *PLoS ONE* **2**, e995. (doi:10.1371/journal.pone.0000995)
15. Caumul R, Polly PD. 2005 Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). *Evolution* **59**, 2460–2472. (doi:10.1111/j.0014-3820.2005.tb00955.x)
16. Viacava P, Baker AM, Blomberg SP, Phillips MJ, Weisbecker V. 2022 Using 3D geometric morphometrics to aid taxonomic and ecological understanding of a recent speciation event within a small Australian marsupial (*Antechinus*: Dasyuridae). *Zool. J. Linn. Soc.* **196**, 963–978. (doi:10.1093/zoolinnean/zlab048)
17. Viacava P, Blomberg SP, Sansalone G, Phillips MJ, Guillerme T, Cameron SF, Wilson RS, Weisbecker V. 2020 Skull shape of a widely distributed, endangered marsupial reveals little evidence of local adaptation between fragmented populations. *Ecol. Evol.* **10**, 9707–9720. (doi:10.1002/ece3.6593)
18. Cardini A, Elton S. 2008 Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. *Biol. J. Linn. Soc.* **93**, 813–834. (doi:10.1111/j.1095-8312.2008.01011.x)
19. Varón-González C, Whelan S, Klingenberg CP. 2020 Estimating phylogenies from shape and similar multidimensional data: why it is not reliable. *Syst. Biol.* **69**, 863–883. (doi:10.1093/sysbio/syaa003)
20. Adams D, Cardini A, Monteiro L, O'Higgins P, Rohlf F. 2011 Morphometrics and phylogenetics: principal components of shape from cranial modules are neither appropriate nor effective cladistic characters. *J. Hum. Evol.* **60**, 240–243. (doi:10.1016/j.jhevol.2010.02.003)
21. Monteiro LR. 2000 Why morphometrics is special: the problem with using partial warps as characters for phylogenetic inference. *Syst. Biol.* **49**, 796–800. (doi:10.1080/106351500750049833)
22. Cardini A, Marco VA. 2022 Procrustes shape cannot be analyzed, interpreted or visualized one landmark at a time. *Evol. Biol.* **49**, 239–254. (doi:10.1007/s11692-022-09565-1)
23. Ascarrunz E, Claude J, Joyce WG. 2019 Estimating the phylogeny of geoemydid turtles (Cryptodira) from landmark data: an assessment of different methods. *PeerJ* **7**, e7476. (doi:10.7717/peerj.7476)
24. Arnaudo ME, Toledo N, Soibelzon L, Bona P. 2019 Phylogenetic signal analysis in the basicranium of Ursidae (Carnivora, Mammalia). *PeerJ* **7**, e6597. (doi:10.7717/peerj.6597)
25. Watanabe A, Slice DE. 2014 The utility of cranial ontogeny for phylogenetic inference: a case study in crocodylians using geometric morphometrics. *J. Evol. Biol.* **27**, 1078–1092. (doi:10.1111/jeb.12382)
26. González-José R, Escapa I, Neves WA, Cúneo R, Pucciarelli HM. 2008 Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. *Nature* **453**, 775–778. (doi:10.1038/nature06891)
27. Catalano SA, Torres A. 2017 Phylogenetic inference based on landmark data in 41 empirical data sets. *Zool. Scripta* **46**, 1–11. (doi:10.1111/zsc.12186)
28. Hetherington AJ, Sherratt E, Ruta M, Wilkinson M, Deline B, Donoghue PC. 2015 Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology* **58**, 393–399. (doi:10.1111/pala.12159)
29. Parins-Fukuchi C. 2018 Bayesian placement of fossils on phylogenies using quantitative morphometric data. *Evolution* **72**, 1801–1814. (doi:10.1111/evo.13516)
30. Smith UE, Hendricks JR. 2013 Geometric morphometric character suites as phylogenetic data: extracting phylogenetic signal from gastropod shells. *Syst. Biol.* **62**, 366–385. (doi:10.1093/sysbio/syt002)
31. Celik MA, Phillips MJ. 2020 Conflict resolution for Mesozoic mammals: reconciling phylogenetic incongruence among anatomical regions. *Front. Genetics* **11**, 651. (doi:10.3389/fgene.2020.00651)
32. Ramírez-Chaves HE, Weisbecker V, Wroe S, Phillips MJ. 2016 Resolving the evolution of the mammalian middle ear using Bayesian inference. *Front. Zool.* **13**, 39. (doi:10.1186/s12983-016-0171-z)
33. Palci A, Lee MS. 2019 Geometric morphometrics, homology and cladistics: review and recommendations. *Cladistics* **35**, 230–242. (doi:10.1111/cla.12340)
34. Cartmill M. 1982 Assessing tarsier affinities: is anatomical description phylogenetically neutral?

Geobios **15**, 279–287. (doi:10.1016/S0016-6995(82)80119-8)

35. Mitchell DR, Kirchhoff CA, Cooke SB, Terhune CE. 2021 Bolstering geometric morphometrics sample sizes with damaged and pathologic specimens: is near enough good enough? *J. Anat.* **238**, 1444–1455. (doi:10.1111/joa.13390)
36. Beck RMD, Voss RS, Jansa SA. 2022 Craniodental morphology and phylogeny of marsupials. *Bull. Am. Mus. Nat. Hist.*, no. 457.
37. O'Leary MA *et al.* 2013 The placental mammal ancestor and the post-K–Pg radiation of placentals. *Science* **339**, 662–667. (doi:10.1126/science.1229237)
38. Turner N. 1848 Observations relating to some of the foramina at the base of the skull in Mammalia, and on the classification of the order Carnivora. *Proc. Zool. Soc. Lond.* **16**, 63–88.
39. Wible JR, Hopson JA. 1993 Basicranial evidence for early mammal phylogeny. In *Mammal phylogeny*, pp. 45–62. Berlin, Germany: Springer.
40. Beer D. 1937 *The development of the vertebrate skull*. Oxford, UK: Clarendon Press.
41. Fleischer G. 1978 *Evolutionary principles of the mammalian middle ear*. Berlin, Germany: Springer.
42. McPhee R. 1981 *Auditory regions of primates and eutherian insectivores: morphology, ontogeny, and character analysis*. Basel, Switzerland: Karger.
43. Luo Z.-X., Kielan-Jaworowska Z, Cifelli RL. 2002 In quest for a phylogeny of Mesozoic mammals. *Acta Palaeontol. Polon.* **47**, 1–78.
44. Gonçalves PR, Christoff AU, Machado LF, Bonvicino CR, Peters FB, Percequillo AR. 2020 Unraveling deep branches of the Sigmodontinae tree (Rodentia: Cricetidae) in eastern South America. *J. Mamm. Evol.* **27**, 139–160. (doi:10.1007/s10914-018-9444-y)
45. Boscaini A, Pujos F, Gaudin TJ. 2019 A reappraisal of the phylogeny of Mylodontidae (Mammalia, Xenarthra) and the divergence of mylodontine and lestodontine sloths. *Zool. Scripta* **48**, 691–710. (doi:10.1111/zsc.12376)
46. Gilbert CC. 2011 Phylogenetic analysis of the African papionin basicranium using 3-D geometric morphometrics: the need for improved methods to account for allometric effects. *Am. J. Phys. Anthrop.* **144**, 60–71. (doi:10.1002/ajpa.21370)
47. Figueirido B, Serrano-Alarcón FJ, Slater GJ, Palmqvist P. 2010 Shape at the cross-roads: homoplasy and history in the evolution of the carnivoran skull towards herbivory. *J. Evol. Biol.* **23**, 2579–2594. (doi:10.1111/j.1420-9101.2010.02117.x)
48. Meloro C, Cáceres NC, Carotenuto F, Sponchiado J, Melo GL, Passaro F, Raia P. 2015 Chewing on the trees: constraints and adaptation in the evolution of the primate mandible. *Evolution* **69**, 1690–1700. (doi:10.1111/evo.12694)
49. Meloro C, Tamagnini D. 2021 Macroevolutionary ecomorphology of the Carnivora skull: adaptations and constraints in the extant species. *Zool. J. Linn. Soc.* **196**, zlab075. (doi:10.1093/zoolinnean/zlab075)
50. Tseng ZJ, Grohé C, Flynn JJ. 2016 A unique feeding strategy of the extinct marine mammal *Kolponomos*: convergence on sabretooths and sea otters. *Proc. R. Soc. B* **283**, 20160044. (doi:10.1098/rspb.2016.0044)
51. Mitchell DR. 2019 The anatomy of a crushing bite: the specialised cranial mechanics of a giant extinct kangaroo. *PLoS ONE* **14**, e0221287. (doi:10.1371/journal.pone.0221287)
52. Horovitz I, Sanchez-Villagra MR. 2003 A morphological analysis of marsupial mammal higher-level phylogenetic relationships. *Cladistics* **19**, 181–212. (doi:10.1111/j.1096-0031.2003.tb00363.x)
53. Rovinsky DS, Evans AR, Adams JW. 2021 Functional ecological convergence between the thylacine and small prey-focused canids. *BMC Ecol. Evol.* **21**, 58. (doi:10.1186/s12862-021-01788-8)
54. 2017–2021 *Materialise Mimics Innovation Suite*, 20.0 edn. See https://www.materialise.com/en/healthcare/mimics-innovation-suite/mimics.
55. Weisbecker V, Speck C, Baker AM. 2020 A tail of evolution: evaluating body length, weight and locomotion as potential drivers of tail length scaling in Australian marsupial mammals. *Zool. J. Linn. Soc.* **188**, 242–254. (doi:10.1093/zoolinnean/zlz055)
56. Heath TA, Hedtke SM, Hillis DM. 2008 Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* **46**, 239–257. (doi:10.3724/SP.J.1002.2008.08016)
57. Ronquist F *et al.* 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
58. Yang Z, Rannala B. 2006 Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **23**, 212–226. (doi:10.1093/molbev/msj024)
59. Fisher DO, Owens IPF, Johnson CN. 2001 The ecological basis of life history variation in marsupials. *Ecology* **82**, 3531–3540. (doi:10.1890/0012-9658(2001)082[3531:TEBOLH]2.0.CO;2)
60. Hume I, Jazwinski E, Flannery T. 1993 Morphology and function of the digestive tract in New Guinean possums. *Aust. J. Zool.* **41**, 85–100. (doi:10.1071/ZO9930085)
61. Voss RS, Jansa SA. 2021 *Opossums: an adaptive radiation of New World marsupials*. Baltimore, MD: Johns Hopkins University Press.
62. Waller PF. 1979 Notes on the trapping and behavior of the Caenolestidae (Marsupialia). *J. Mammal.* **60**, 390–395. (doi:10.2307/1379811)
63. Stratovan Corporation. 2019 *Stratovan Checkpoint. 3D shape analysis & morphometrics software*, 2019.03.04.1102 edn. See https://www.stratovan.com/products/checkpoint.
64. Sánchez-Villagra MR, Wible JR. 2002 Patterns of evolutionary transformation in the petrosal bone and some basicranial features in marsupial mammals, with special reference to didelphids. *J. Zool. Syst. Evol. Res.* **40**, 26–45. (doi:10.1046/j.1439-0469.2002.00173.x)
65. Baken EK, Collyer ML, Kaliontzopoulou A, Adams DC. 2021 geomorph v4. 0 and gmShiny: enhanced analytics and a new graphical interface for a comprehensive morphometric experience. *Methods Ecol. Evol.* **12**, 2355–2363. (doi:10.1111/2041-210X.13723)
66. Adams DC, Collyer M, Kaliontzopoulou A, Sherratt E. 2016 *geomorph: Software for geometric morphometric analyses*. See https://github.com/geomorphR/geomorph.
67. Marcy AE, Guillerme T, Sherratt E, Rowe KC, Phillips MJ, Weisbecker V. 2020 Australian rodents reveal conserved cranial evolutionary allometry across 10 million years of murid evolution. *Am. Nat.* **196**, 755–768. (doi:10.1086/711398)
68. Sansom RS, Wills MA, Williams T. 2017 Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Syst. Biol.* **66**, 813–822. (doi:10.1093/sysbio/syw116)
69. Kraatz BP, Sherratt E, Bumacod N, Wedel MJ. 2015 Ecological correlates to cranial morphology in leporids (Mammalia, Lagomorpha). *PeerJ* **3**, e844. (doi:10.7717/peerj.844)
70. Clavel J, Escarguel G, Merceron G. 2015 mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319. (doi:10.1111/2041-210X.12420)
71. Clavel J, Morlon H. 2020 Reliable phylogenetic regressions for multivariate comparative data: illustration with the MANOVA and application to the effect of diet on mandible morphology in phyllostomid bats. *Syst. Biol.* **69**, 927–943. (doi:10.1093/sysbio/syaa010)
72. Clavel J, Morlon H. 2020 Data from: Reliable phylogenetic regressions for multivariate comparative data: illustration with the MANOVA and application to the effect of diet on mandible morphology in phyllostomid bats. Dryad Digital Repository. (doi:10.5061/dryad.jsxksn052)
73. Uyeda JC, Zenil-Ferguson R, Pennell MW. 2018 Rethinking phylogenetic comparative methods. *Syst. Biol.* **67**, 1091–1109. (doi:10.1093/sysbio/syy031)
74. Adams DC. 2014 A generalized *K* statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* **63**, 685–697. (doi:10.1093/sysbio/syu030)
75. Guillerme T, Puttick MN, Marcy AE, Weisbecker V. 2020 Shifting spaces: which disparity or dissimilarity measurement best summarize occupancy in multidimensional spaces? *Ecol. Evol.* **10**, 7261–7275. (doi:10.1002/ece3.6452)
76. Collyer ML, Adams DC. 2021 Phylogenetically aligned component analysis. *Methods Ecol. Evol.* **12**, 13515. (doi:10.1111/2041-210X.13515)
77. Schliep K, Potts A, Morrison D, Grimm G. 2017 Intertwining phylogenetic trees and networks. *Methods Ecol. Evol.* **8**, 1212–1220. (doi:10.1111/2041-210X.12760)
78. Goloboff PA, Catalano SA. 2016 TNT version 1.5, including a full implementation of phylogenetic

morphometrics. *Cladistics* **32**, 221–238. (doi:10.1111/cla.12160)

79. Catalano SA, Goloboff PA, Giannini NP. 2010 Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework. *Cladistics* **26**, 539–549. (doi:10.1111/j.1096-0031.2010.00302.x)
80. Swofford DL. 2003 *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*, 4th edn. Sunderland, UK: Sinauer Associates.
81. Smith MR. 2020 Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. *Bioinformatics* **36**, 5007–5013. (doi:10.1093/bioinformatics/btaa614)
82. Oksanen J *et al.* 2022 *vegan: Community Ecology Package*, 2.6-2 edn. See https://github.com/vegandevs/vegan.
83. Swofford D, Olsen G, Waddell P, Hillis D. 1996 Phylogeny reconstruction. In *Molecular systematics*, 2nd edn (eds DM Hillis, C Moritz, BK Mable), pp. 407–514. Sunderland, MA: Sinauer Associates.
84. Wible J, Rougier G, Novacek M. 2005 Anatomical evidence for superordinal/ordinal eutherian taxa in the Cretaceous. In *The rise of placental mammals: origins and relationships of the major extant clades* (eds KD Rose, JD Archibald), pp. 15–36. Baltimore, MD: Johns Hopkins University Press.
85. Chemisquy MA, Tarquini SD, Romano Muñoz CO, Prevosti FJ. 2021 Form, function and evolution of the skull of didelphid marsupials (Didelphimorphia: Didelphidae). *J. Mamm. Evol.* **28**, 23–33. (doi:10.1007/s10914-019-09495-4)
86. Alvarez A, Ercoli MD, Prevosti FJ. 2013 Locomotion in some small to medium-sized mammals: a geometric morphometric analysis of the penultimate lumbar vertebra, pelvis and hindlimbs. *Zoology* **116**, 356–371. (doi:10.1016/j.zool.2013.08.007)
87. Dumont M, Wall CE, Botton-Divet L, Goswami A, Peigné S, Fabre A-C. 2016 Do functional demands associated with locomotor habitat, diet, and activity pattern drive skull shape evolution in musteloid carnivorans? *Biol. J. Linn. Soc.* **117**, 858–878. (doi:10.1111/bij.12719)
88. Hautier L, Lebrun R, Cox PG. 2012 Patterns of covariation in the masticatory apparatus of hystricognathous rodents: implications for evolution and diversification. *J. Morph.* **273**, 1319–1337. (doi:10.1002/jmor.20061)
89. Melstrom KM, Angielczyk KD, Ritterbush KA, Irmis RB. 2021 The limits of convergence: the roles of phylogeny and dietary ecology in shaping non-avian amniote crania. *R. Soc. Open Sci.* **8**, 202145. (doi:10.1098/rsos.202145)
90. Breidbach O, Jost J. 2006 On the gestalt concept. *Theor. Biosci.* **125**, 19–36. (doi:10.1016/j.thbio.2006.02.001)
91. Phillips MJ, Cascini M, Celik M. 2022 Identifying complex DNA contamination in pig-footed bandicoots helps to clarify an anomalous ecological transition. *Diversity* **14**, 352. (doi:10.3390/d14050352)
92. Goswami A, Smaers JB, Soligo C, Polly PD. 2014 The macroevolutionary consequences of phenotypic integration: from development to deep time. *Phil. Trans. R. Soc. B* **369**, 20130254. (doi:10.1098/rstb.2013.0254).
93. Fabre A-C, Dowling C, Portela Miguez R, Fernandez V, Noirault E, Goswami A. 2021 Functional constraints during development limit jaw shape evolution in marsupials. *Proc. R. Soc. B* **288**, 20210319. (doi:10.1098/rspb.2021.0319)
94. Tamagnini D, Meloro C, Raia P, Maiorano L. 2021 Testing the occurrence of convergence in the craniomandibular shape evolution of living carnivorans. *Evolution* **75**, 1738–1752. (doi:10.1111/evo.14229)
95. Meloro C, Clauss M, Raia P. 2015 Ecomorphology of Carnivora challenges convergent evolution. *Organisms Div. Evol.* **15**, 711–720. (doi:10.1007/s13127-015-0227-5)
96. Cardini A, Elton S, Kovarovic K, Strand Viðarsdóttir U, Polly PD. 2021 On the misidentification of species: sampling error in primates and other mammals using geometric morphometrics in more than 4000 individuals. *Evol. Biol.* **48**, 190–220. (doi:10.1007/s11692-021-09531-3)
97. Kamilar JM, Cooper N. 2013 Phylogenetic signal in primate behaviour, ecology and life history. *Phil. Trans. R. Soc. B* **368**, 20120341. (doi:10.1098/rstb.2012.0341)
98. Klingenberg CP, Gidaszewski NA. 2010 Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Syst. Biol.* **59**, 245–261. (doi:10.1093/sysbio/syp106)
99. Meloro C, Raia P, Carotenuto F, Cobb SN. 2011 Phylogenetic signal, function and integration in the subunits of the carnivoran mandible. *Evol. Biol.* **38**, 465–475. (doi:10.1007/s11692-011-9135-6)
100. Álvarez A, Perez SI, Verzi DH. 2011 Ecological and phylogenetic influence on mandible shape variation of South American caviomorph rodents (Rodentia: Hystricomorpha). *Biol. J. Linn. Soc.* **102**, 828–837. (doi:10.1111/j.1095-8312.2011.01622.x)
101. Álvarez A, Ercoli MD, Verzi DH. 2019 Integration and diversity of the caviomorph mandible (Rodentia: Hystricomorpha): assessing the evolutionary history through fossils and ancestral shape reconstructions. *Zool. J. Linn. Soc.* **188**, 276–301. (doi:10.1093/zoolinnean/zlz071)
102. Morgan CC. 2009 Geometric morphometrics of the scapula of South American caviomorph rodents (Rodentia: Hystricognathi): form, function and phylogeny. *Mamm. Biol.* **74**, 497–506. (doi:10.1016/j.mambio.2008.09.006)
103. Weisbecker V, Beck RMD, Guillerme T, Harrington AR, Lange-Hodgson L, Lee MSY, Mardon K, Phillips MJ. 2023 Multiple modes of inference reveal less phylogenetic signal in marsupial basicranial shape compared with the rest of the cranium. Figshare. (doi:10.6084/m9.figshare.c.6566745)