Geometric morphometrics out-performs linear-based methods in the taxonomic resolution of a mammalian species complex

Pietro Viacava1,2?, Simone Blomberg1, Vera Weisbecker2

1School of Biological Sciences, The University of Queensland, St Lucia, Queensland, Australia

2College of Science and Engineering, Flinders University, Adelaide, South Australia, Australia

Abstract

Morphology-based taxonomic research has predominantly used linear morphometrics (LMM) to quantitatively measure skulls to distinguish species. The choice of which measurements to collect generally relies on the expertise of the investigators or a set of standard measurements, but this practice may ignore less obvious or common discriminatory characters. In addition, taxonomic analyses often ignore allometry and thus the potential for subgroups of an otherwise cohesive population to differ in shape purely due to size differences. 3D geometric morphometrics (GMM) is more complicated as an acquisition technique, but can offer a more holistic characterization of shape and provides a rigorous toolkit for allometry assessments. In this study, I used linear discriminant analysis to assess the discriminatory performance of four published LMM protocols and a 3D GMM dataset for three clades of antechinus known to differ subtly in shape. We assessed discrimination of raw data (which are frequently used by taxonomists); after removal of isometry; and after allometric correction. We found that group discrimination among raw data was high for LMM, but GMM produced better results in group discrimination after the size and allometry treatments. High measurement redundancy in LMM protocols can result in relatively high allometry but low discriminatory performance. These findings suggest that taxonomic measurement protocols might benefit from GMM-based pilot studies, because this offers the option of differentiating allometric and non-allometric shape differences between species, which can then benefit the development of the easier-to-apply LMM protocols. (CITE LIONEL HAUTIER, IN A CONTEXT OF RODENTS NOT FOR TAXONOMIC PURPOSES BUT WORTH CITING HIS WORK BECAUSE IT SHOWS THAT IT WORKS AND THAT LMM AND GMM CAN IMPROVE EACH OTHER)

Introduction

Morphometric measurements are an important tool in efforts to differentiate mammalian species from each other. Measurements of an animal’s skull provide important data which can be used in the delimitation of species or Evolutionary Significant Units (ESUs) (Viacava et al., 2020, 2021). (DESCRIBE HOW MORPHOMETRICS IS AN IMPORTANT PART OF SPECIES DELIMITATION, SUPPORTING GENETICS BUT ALSO CRUCIAL FOR UNDERSTANDING HOW GENETIC AND MORPHOLOGICAL DIFFERENTIATION CORRESPOND IN CLOSE RELATIVES).

Morphometrics-based taxonomic differentiation remains mostly the domain of linear morphometrics (LMM) (REFS), mainly because linear measurements are easily taken and comparable to results from past studies. However, this approach has several limitations. Typically, taxonomists choose a set of linear distances based on their expertise of the morphology of the taxon in question. This could pose a problem when these linear protocols are not standardized among morphometricians, obtaining potentially different acquisitions of data in studies of similar taxa but based on different protocols. In addition, the linear distances measured in taxonomic diagnoses often include maximum and minimum heights, widths and lengths that are easily identifiable to the eye. However, these measurements may characterize conditions that are not necessarily biologically homologous across taxa. This is because, when shapes differ, maximum or minimum distances may not be comparable among individuals, as the point-to-point distance may relate to different reference points that are not necessarily homologous (Figure 1). Additionally, LMM describes the distance between two points, missing potentially relevant variation between shapes and not retaining information about the original shape. Lastly, size differences are not often accounted for in taxonomic studies using LMM, but there is a problem of redundancy when linear measurements contain within them other linear measurements (e.g., multiple measurements along the longitudinal axis of the skull). Similarly, the frequent use of proportional ratios (e.g. skull length vs. width) is problematic because many species display intraspecific allometry, such that genetically similar individuals will differ in a ratio simply because they differ in size (Sidlauskas et al., 2011).

A potential refinement of LMM protocols could be offered by the use of Geometric Morphometrics (GMM). GMM uses the coordinates of anatomical reference points as identifiable shape variables in most or all the specimens in a given dataset. Since the 1990s, this technique (Adams et al., 2004) has become the standard method for shape characterization in evolutionary and ecological morphometrics and has a very mature analytical toolkit (Fruciano, 2016). The holistic characterization of the biological specimens under study and the graphical output of the associated shape changes provide a powerful and alternative tool for biological inferences in many contexts (Klingenberg, 2016; Stone, 1997), a definite advantage that linear morphometrics cannot offer.

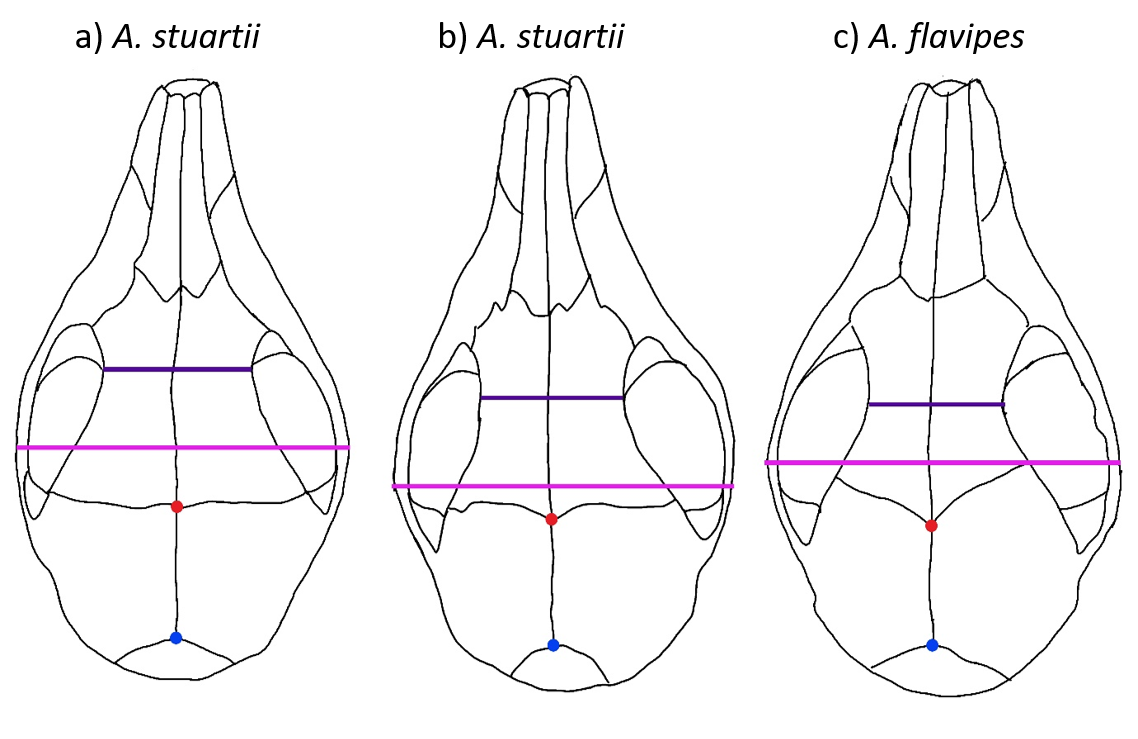


Figure 1: Three specimens (CM12785, CM6540 and CM10548) outlining two commonly used linear distances: the width of greatest constriction of orbitotemporal fossa (dark purple) and the maximum width of cranium measured across zygomatic arches (pink). In addition, two type I homologous landmarks (by suture intersection) are depicted: the fronto-parietal suture in midline (red dot) an the parietal-interparietal suture in midline (blue dot). The two examples of maximum and minimum distances are measured at different anatomical positions relative to the homologous landmarks and other sutures in the skull, indicating a possibly serious lack of homology.

GMM is rarely used in taxonomic studies, probably because linear measurements are easier to acquire and analyse, and because there is a large body of literature on taxonomic measurements that has been widely used for centuries (FIND SOME REFS). In addition, geometric morphometric data acquisition can be more complex, requiring digitisation of either photographs or 3D specimen representations, which generally involve specialized equipment. The statistical analyses required, while very well developed and versatile (REFS), are also specialised and involve high-dimensional data, Therefore, they may not be perceived to be straightforward as the statistical toolkits used in LMM analyses.

We argue that GMM analyses are a useful addition to taxonomic studies because the technique addresses several of the issues of LMM raised above. XXXXXXX

HERE EXPLAIN HOW GMM IMPROVES ON THE ISSUES OF LMM IN THE ORDER OF THE 2ND INTRO PARAGRAPH –

FIRST, LINEAR DISTANCES BASED ON EXPERTISE RATHER THAN A BLANKET ALL-ENCOMPASSING PROTOCOL (AND MENTION HAT GMM PROTOCOLS NEED TO CAREFULLY CONSIDER COVERAGE)

SECOND, NON-HOMOLOGOUS MAX/MIN MEASUREMENTS

THIRDS, DISTANCE BETWEEN TWO POINTS VS ALL POINT RELATIVE TO EACH OTHER

THESE DON’T HAVE TO BE AS EXPANSIVE AS THE SIZE ONE, WHICH IS OF COURSE MOST IMPORTANT, BUT IF I SET UP THE LMM DISADVANTAGES ABOVE I NEED TO OUTLINE HOW GMM ADDRESSES THEM LATER

One of these is the explicit treatment of size. The use of Procrustes superimposition allows the removal of the size component from the dataset by scaling all specimens to the same size. This procedure results in two components: a proxy for size called centroid size and a multivariate shape component (Kendall, 1989). These can then be used for analyses of allometry (shape changes disproportionate to size) in the form of a shape vs size regression (Klingenberg, 2016). This substantially improves on the issue of accounting for isometric and allometric variation, which can have serious implications for taxon delimitation. Allometric effects in particular can give an impression of species differentiation, when cranial allometry is generally present within most mammalian species (REFS, PIETRO, ARIEL AND OTHERS) and may not be related to morphological divergence due to a speciation event (Sidlauskas et al., 2011). However, such allometric variation has been regarded as irrelevant in taxonomy because if shape differences were strictly due to size differences then they are likely to be the differences between small and large animals within a taxonomic group (Pilbeam & Gould, 1974; Seifert, 2008; Wood & Stack, 1980). In contrast, non-allometric shape changes are thought to be caused by independent adaptive processes (Huxley, 1931; Gould, 1975). Thus, it is recommended to include allometric analyses in taxonomic studies in order to properly delimit species (Outomuro & Johansson, 2017; Sidlauskas et al., 2011). However, even in a case where taxonomic differentiation is driven purely by selection for size and coincides only with allometry effects, this represents important information on the differentiation process and should be considered.

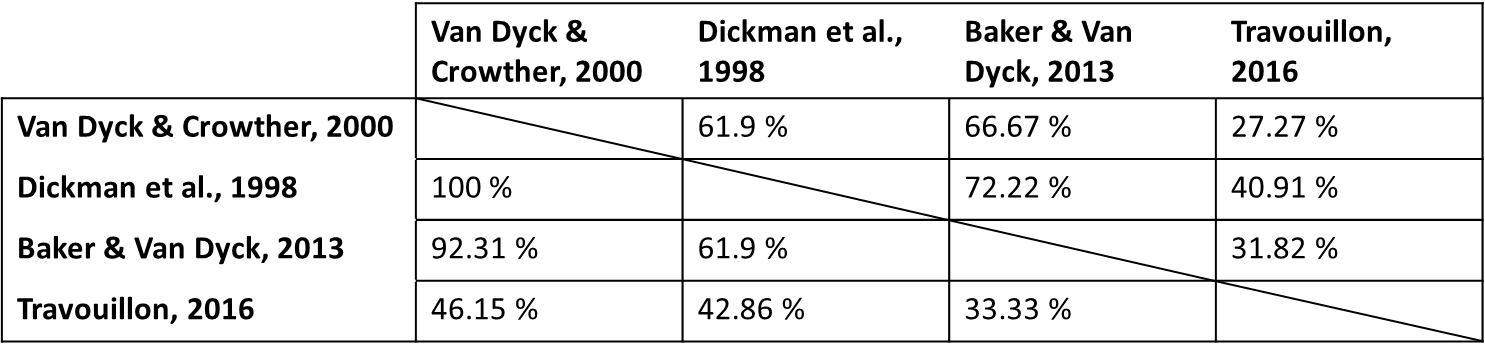
Here, we compare the taxonomic differentiation performance of conventional, linear-based morphometrics and 3D geometric morphometrics in a species complex that includes three genetically differentiated taxonomic groups containing subtle morphological differences (Viacava et al., 2021), the *A. stuartii* / *A. subtropicus* species complex. This represents a scenario where finer-grained differences (e.g., allometric effects or differences in non-homologous measurements) need to be identified and interpreted with great care to understand their pertinence to group differentiation. The taxonomic situation of this species complex is also useful because three linear morphometric protocols have been used pertaining to the genus, allowing an assessment of how important protocol choice can be to the delimitation of taxonomic units. We add to this also a more generic protocol developed for bandicoots (Travouillon, 2016), with a particularly high coverage of linear measurements. We use linear discriminant analysis to ask how well the four protocols and our GMM protocol perform in a context typical to taxonomy (without consideration of size, or with consideration only of isometry) and compare it to the analysis pipeline typically taken I studies of GMM (with allometric variation accounted for as well).

Material and methods

All analyses are based on a 3D landmark coordinate dataset from Viacava et al. (2021), which includes high-coerage 3D landmarked crania with 412 landmarks (82 fixed landmarks, 185 curved semilandmarks and 145 surface semilandmarks) of 136 adult individuals reconstructed in virtual 3D images. All analyses were performed in R version 4.0.4 (R Core Team, 2021). The code (?) and raw data are available on (GITHUB ADDRESS).

We identified four linear measurement protocols that represent traditional morphometric methods commonly used in Australian mammal taxonomy, but are also specific to *Antechinus*. These include a protocol used for a species contained in the species complex studied here, *A. subtropicus* (Van Dyck & Crowther, 2000), a sister species of the species complex studied here, *A. agilis* (Dickman et al., 1998), a species within the genus *Antechinus*, *A. flavipes* (Baker & Van Dyck, 2013), and a comprehensive protocol that was developed for Peramelemorphs (bandicoots) (Travouillon, 2016). The last protocol is not necessarily expected to apply well to the genus *Antechinus* because it was designed for a different order of marsupials. However, it was chosen as a useful comparison of performance with the other three sets of linear measurements, representing one of the most comprehensive protocols in the morphometric study of Australian mammals. All of these protocols differ from each other, but overlap in some measurements (Table 1).

Table 1. Degree of overlap of linear measurements between protocols. The LMM protocols in the rows cover a fraction of the LMM protocols in the columns.



To obtain the linear measurement data, we extracted the linear distances of each protocol that could be estimated most appropriately from the coordinates of the landmarks used for the geometric morphometric approach (Supplementary Table S4.1). These measurements were not exactly the same as calliper measurements; however, we assume that slight inconsistencies between linear-based and 3D landmark-based distances are acceptable because they were taken in a consistent fashion and the representation of shape taken with the linear distances is not lost. We averaged right and left measurements whenever possible.

### Isometry and allometry

In geometric morphometrics, the isometric component of shape (i.e. the shape that varies in a 1:1 proportion with size) is generally removed from the dataset through the scaling procedure of the Procrustes superimposition. This step brings all specimens to the same size, producing “isometry-free” shape coordinates and a centroid size (Dryden & Mardia, 2016; Klingenberg, 2016) for each specimen. Centroid size can be used subsequently as a proxy for specimen size. To approximate this effect in the LMM context, we used an approach that is analogous to centroid size extraction by deriving the geometric mean of all variables as the centroid size, and using log-shape ratios [log10(measurement/geometric mean)] as isometry-free shape variables. This ensures that each dataset can be analysed in an approximately equivalent way (Claude, 2013; Mosimann, 1970).

There is ample evidence in GMM that patterns of allometry can seriously impact the interpretation of shape variation (Cardini & Polly, 2013; Kaliontzopoulou et al., 2008; Outomuro & Johansson, 2017; Seifert, 2008; Yazdi, 2014). In the context of taxonomic differentiation between species, it seems particularly important to assess whether two groups differ in shape because of allometry: if one population has different proportions compared to the other purely because its individuals are larger, this might not constitute a case for diagnosing two groups as species (INTEGRATE THIS PART OF THE PARAGRAPH INTO INTRODUCTION). In order to assess the effect of allometry, we regressed the Procrustes shape variation vs log(centroid size) with the “geomorph” (Adams et al., 2016) function “procD.lm”. For LMM, we regressed the linear data vs log(geometric mean) with the “lm.rrpp” function of the “RRPP” package (Collyer & Adams, 2018). We considered both centroid size and geometric mean as proxies for size in the context of geometric and linear morphometrics. We also computed “allometry-free” datasets for the classification analyses below, by using the residuals from the allometric regressions. In summary, three types of morphological data were obtained and analysed for the LMM protocols and the GMM dataset: a) raw 3D coordinates obtained from a partial Procrustes superimposition (GMM) and raw linear measurements (LMM), b) shape after Procrustes superimposition (GMM) and log-shape ratios as explained above (LMM), and c) allometry-corrected shape for both. In the case of “raw shape”, this type of data is typically called “form” in geometric morphometrics (shape plus size). However, for practical purposes, we will further call the types of morphological data explained above as “raw”, “isometry-free” and “allometry-free” shape, respectively. Allometric regressions were performed with 1000 permutations and p-values were calculated using Goodall’s F-test (REF).

### Ordination

I computed Principal Component Analysis (PCA) for each treatment (raw, isometry-free and allometry-free measurements) and each linear measurement protocol and geometric morphometrics. By labelling the groups (clades) in each PCA plot of the first two PCs, it is possible to observe how group discrimination changes within the main variation after scaling and allometric correction. However, note that lack of differentiation of groups in PC1/PC2 space does not mean that the groups are not differentiated; PCA is agnostic to groupings, such that variation that differentiates a particular group can also be “smeared” across many Principal components (Bookstein, 2015, 2017a, 2017b; Klingenberg et al., 1996; Weisbecker et al., 2019).

### Classification rule

To assess how well specimens are predicted to belong to each group based on the different morphological data, I calculated classification performance measures based on the results of a Linear Discriminant Analysis. I then performed a Linear Discriminant Analysis with the clade identity as a group factor and provided an equal prior on class membership to the three groups. I plotted the two linear discriminants for each treatment (raw, isometry-free and allometry-free measurements), and for each linear measurement protocol and the geometric morphometrics protocol. Next, I used a machine learning model known as leave-one-out cross validation procedure to calculate the posterior values (Venables & Ripley, 2002). These posterior values allowed us to calculate Garczarek’s classification performance measures (Garczarek & Weihs, 2003) with the ‘klaR’ package for R (Weihs et al., 2005) in the form of Correctness Rate (CR), Accuracy (AC), Ability to Separate (AS), Confidence (CF), and Confidence for each class. The CR and AC values estimate the degree of validity (“quality”) of the linear discriminant analysis from the predicted values based on the true values. AS corresponds to the distance between the posterior values and the assigned groups and CF measures the degree of confidence to which the groups have been assigned – both AS and CF estimate the “certainty” of the result of the linear discriminant analysis (Dr. Karsten Luebke pers. comm.).

Results

### Allometry

Allometric regressions were performed with 1000 permutations and p-values were calculated using a Goodall’s F-test. All LMM and GMM protocols were significantly allometric – influence of size on isometry-free shape (Table 2). The least allometry was detected when using Van Dyck & Crowther’s linear protocol (R2 = 0.079, p = 0.001). Conversely, the most allometry was found when using Travouillon’s linear measurements (R2 = 0.251, p = 0.001). Dickman et al., Baker & Van Dyck and GMM (R2 ≈ 0.144, 0.113 and 0.132; p=0.001 for all three) identified a similar allometric effect (Table 2).

### Ordination

The first Principal Component (PC1) of the three LMM protocols used in antechinuses – all except Travouillon – accounted for more than 70% of morphological variation in all cases: raw, isometry-free and allometry-free (Figure 2). Only Travouillon et al.’s linear distances had a considerable effect on the morphological variation accounted for by PC1 after the removal of isometry, dropping from 73.36% to 38.33%, and to 24.65% after allometric correction. I also observed a reduction in morphological variation accounted by PC1 after removal of isometry in the GMM protocol, from 78.77% to 19.43%, and a slight decrease after allometric correction to 14.78%. Grouping of the three clades was affected in all three stages of data treatment; in all cases, isometry removal contributed to the assemblage of the groups along PC1 and allometric correction scattered the three groups showing unclear grouping (Figure 2).

### Classification rule

The Linear Discriminant Analysis showed similar grouping of clades for raw and isometry-free measurements. For the GMM protocol, the removal of isometry had increased group differentiation (Figure 3). The LMM protocols showed a considerable decrease in group differentiation after removal of allometry (Figure 3 and Table 2). For GMM, the removal of allometry did not affect group differentiation as much as for LMM (Figure 3 and Table 2).

For raw data, the classification performance measures were reasonably high in all four LMM protocols (Table 2). After isometry removal and allometric correction, these measures decreased at varying degrees for all LMM protocols. GMM performed relatively better than LMM at group discrimination at the raw data stage. After the removal of isometry, GMM performed similarly to LMM protocols in CR and AC (“quality” measures) and better in AS and CF (“certainty” measures). After allometric correction, a large decrease in CR and AC was observed in GMM data despite similar performance in AS and CF.

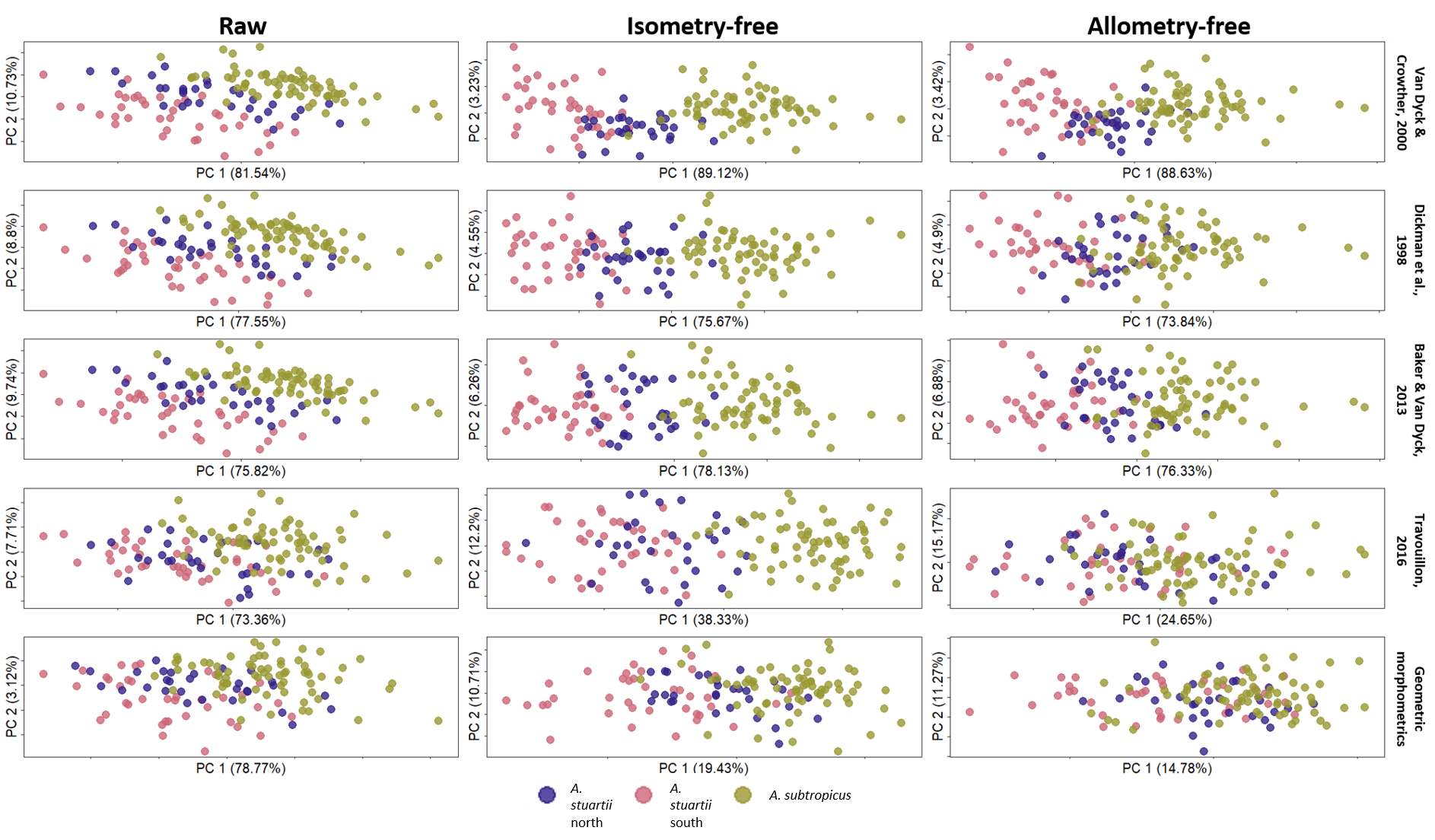


Figure 2: Principal Component Analyses plot for all raw, isometry-free and allometry-free datasets. These include the four linear measurement protocols and the geometric morphometrics approach. Only the first two Principal Components are shown.

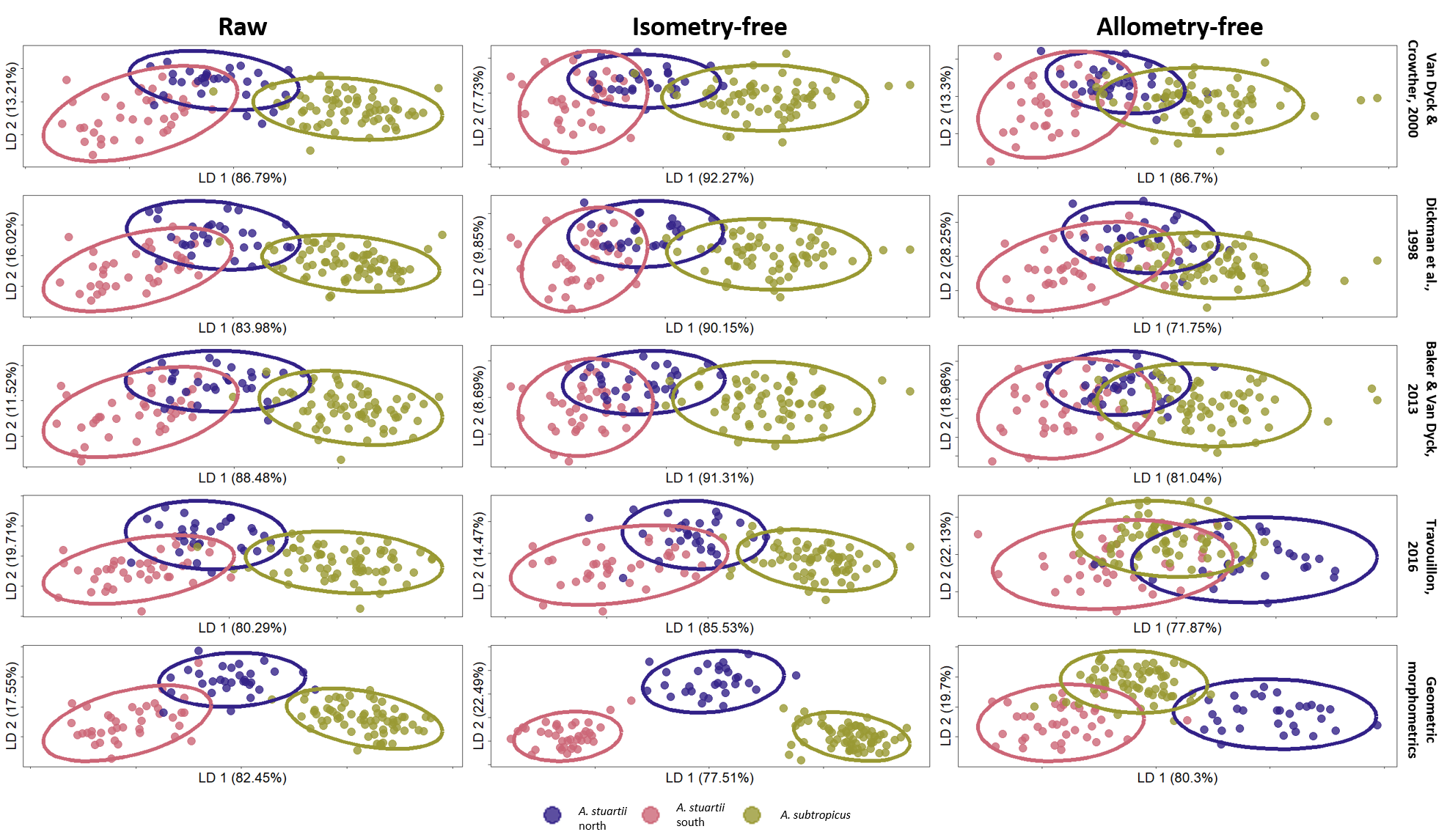


Figure 3: Linear Discriminant Analyses plot for all raw, isometry-free and allometry-free datasets used in this study. These include the four linear measurement protocols and the geometric morphometrics approach. Ellipses were computed at 95%condifence intervals.

Table 2: Classification performance measures (Garczarek, 2002) of the four linear measurement protocols and geometric morphometrics. For each of them, the classification performance measures were computed with raw datasets, after size treatment, and after allometry correction. Allometric regression results are also indicated in the last row.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Van Dyck & Crowther, 2000** | | | **Dickman et al., 1998** | | | **Baker & Van Dyck, 2013** | | | **Travouillon, 2016** | | | **GMM** | | |
| **Raw** | **Size-free** | **Allometry-free** | **Raw** | **Size-free** | **Allometry-free** | **Raw** | **Size-free** | **Allometry-free** | **Raw** | **Size-free** | **Allometry-free** | **Raw** | **Size-free** | **Allometry-free** |
| **Correctness Rate** | 0.904 | 0.875 | 0.846 | 0.904 | 0.853 | 0.757 | 0.875 | 0.86 | 0.757 | 0.882 | 0.853 | 0.684 | 0.926 | 0.86 | 0.581 |
| **Accuracy** | 0.759 | 0.686 | 0.553 | 0.779 | 0.681 | 0.517 | 0.741 | 0.687 | 0.503 | 0.764 | 0.712 | 0.335 | 0.843 | 0.761 | 0.286 |
| **Ability to Separate** | 0.855 | 0.763 | 0.655 | 0.87 | 0.784 | 0.706 | 0.837 | 0.793 | 0.666 | 0.879 | 0.858 | 0.605 | 0.943 | 0.983 | 0.879 |
| **Confidence** | 0.915 | 861 | 0.791 | 0.924 | 0.874 | 0.817 | 0.905 | 0.879 | 0.794 | 0.927 | 0.914 | 0.754 | 0.967 | 0.99 | 0.928 |
| **Confidence for each true class** | North:  0.89 | North:  0.773 | North:  0.72 | North:  0.889 | North:  0.802 | North:  0.794 | North:  0.86 | North:  0.776 | North:  0.724 | North:  0.892 | North:  0.863 | North:  0.81 | North:  0.963 | North:  0.961 | North:  0.947 |
| South:  0.875 | South:  0.807 | South:  0.786 | South:  0.886 | South:  0.817 | South:  0.833 | South:  0.87 | South:  0.819 | South:  0.781 | South:  0.926 | South:  0.897 | South:  0.757 | South:  0.953 | South:  0.997 | South:  0.944 |
| Sub:  0.949 | Sub:  0.931 | Sub:  0.824 | Sub:  0.961 | Sub:  0.937 | Sub:  0.818 | Sub:  0.945 | Sub:  0.958 | Sub:  0.832 | Sub:  0.944 | Sub:  0.946 | Sub:  0.728 | Sub:  0.976 | Sub:  0.998 | Sub:  0.91 |
| **Allometry** | R² = 0.079, F = 11.475, p = 0.001 | | | R² = 0.144, F = 22.48, p = 0.001 | | | R² = 0.113, F = 17.064, p = 0.001 | | | R² = 0.251, F = 44.99, p = 0.001 | | | R² = 0.132, F = 20.403, p = 0.001 | | |

Discussion

I expected GMM to improve the group discrimination performance relative to LMM because of its holistic shape characterization. However, if the number of variables (in these cases, linear distances and 3D coordinates of landmarks) discriminating groups was proportionally higher, as could be the case in LMM relative to GMM (Mitteroecker & Gunz, 2009; Slice, 2007), then I also expected that the LMM protocols that included the highest differentiable characters would perform correctly in classification measures. I found that GMM performed slightly better when considering raw and isometry-free observations. However, the performance of LMM varied chiefly according to the consideration of the few linear distances that had discriminant power in the LMM protocols. In other words, the traditional morphometric protocols were highly dependent on the choice of the measurements. The fewer variables relative to GMM may superficially aid the LMM protocols in group discrimination if the selected linear distances are the “real” best discriminatory ones. In the case of 3D GMM, this dependence on measurement choice is expected to be less pronounced if the creation of the landmarking template relies on the agnostic nature of recognizable homologous reference points present in all specimens in a given dataset. The selection of the variables, or landmarks (coordinates), involves the construction of a template that attempts the optimal anatomical coverage with diverse homologous points. This process does not necessarily focus on the most variable regions because it can rely on finding those differentiable shape patterns at the analytical step (Webster & Sheets, 2010). Thus, proportionally, some linear measurement protocols might be best at discriminating but this will only be the case if the linear distances selected are best at discriminating in “reality”. To illustrate this, Travouillon’s protocol (Travouillon, 2016) provides an example. The relatively low classification performance of this protocol is almost certainly due to the choice of the linear distances, which were not selected for antechinus morphological characterization, contrary to the other three LMM protocols.

Intriguingly, the biplot of the main variation (PC1 and PC2) resulting from the PCA of the GMM data revealed unclear grouping of the clades (see Figure 2). Thus, an interpretation solely based on the first principal components can be misleading (Schreiber, 2021). Indeed, the PCA biplot based on LMM reflected a clearer grouping of clades relative to GMM. Superficially, this suggests that LMM may distinguish better clades based on linear distances. This outcome is chiefly due to the lower dimensionality of the dataset and the fact that, in these particular cases, the linear distances chosen largely reflected the differences among clades. However, this consideration can be deceptive given the low variance explained by the first principal component in GMM, where the morphological variance tends to be distributed along a larger number of principal components. This simply reflects the fact that the GMM dataset contains far more variation overall, much of which does not differentiate clades. However, some principal components important for group discrimination may also remain hidden in the lower ranked ones (Bookstein, 2017b; Klingenberg et al., 1996; Weisbecker et al., 2019). The relevance of ignored morphological variance in a PCA biplot in GMM is revealed in the LDA where 95% of the variance was taken into account, showing a similar contribution to group discrimination in GMM and LMM. Moreover, the classification performance measures further reflect how the ability of GMM to differentiate among clades performs exceedingly well (see Table 2).

In the isometry removal step (between raw and isometry-free datasets), decreases in the “certainty” measures of classification in LMM datasets contrasted with increases in the GMM protocol. This can also be observed in the LDA biplot in Figure 3, where group separation in GMM is clearly visible in the isometry-free dataset. In this aspect of isometry removal, the contribution of geometric morphometrics towards isometry-free group separation may be a substantial improvement in the way we regard size and shape as independent variables for subsequent allometric analyses. In the case of geometric morphometrics, the larI age number of landmarks may contribute to a holistic characterization of size – in the form of centroid size (Mitteroecker & Gunz, 2009). In the LMM context, the linear distances may contain fewer aspects of the size of the skull (Farkas et al., 2002; Slice, 2006). For example, if we measured only the length or the width of a skull, other linear distances associated with size-related shape could be ignored, such as the width of the snout. This can be a problem because it disregards measures that are characterizing the size of a three-dimensional object (Adams et al., 2004). Furthermore, if size is not characterized well, further consequences on the independence of a size and shape variable can undermine allometric analyses in the form of a size vs shape regression (Klingenberg, 2016).

The removal of shape variation due to allometry (the step from “isometry-free” to “allometry-free”) had a larger decrease in the classification performance measures than the previous step of removal of isometry (from “raw” to “isometry-free”). This meant that some of the shape variation discriminating groups was correlated to size variation. However, this removal of allometry had a greater effect on classification performance in GMM, as demonstrated by the large decline in Correctness Rate and Accuracy (see Table 2). This could be either because of redundancy in the information of nearby landmarks and semilandmarks, or because GMM deals more effectively with allometry. With Travouillon’s linear measurements (Travouillon, 2016), removal of allometry has a similar effect on Accuracy despite the particularly stronger allometric relationship captured by Travouillon’s linear measurements. This “large amount of allometry” captured by Travouillon may be caused by the redundancy of some linear measurements that may exacerbate some shape patterns driven by size (e.g., because most measurements are along the length of the skull and therefore contain a component of skull length). If this was the case in GMM, the Ability to Separate, reduced drastically in Travouillon, would also be reduced in GMM. However, we observed high performance in Ability to Separate in GMM after the removal of allometry. The contrasting results for AS and AC in GMM show a more thorough way of considering allometry-driven shape patterns relative to less informative linear measurements. However, it may also be an indicator of a linear discriminant analysis wrongly assigning classes with false “certainty” due to the poor variables (PC scores) / observations (number of individuals) ratio typically encountered in GMM. Note also that the relatively low performance of Travouillon’s protocol may be particularly unsuitable for dasyuromorphs because it was created with view to species with a very different skull shape, the peramelids (bandicoots). This may be the reason why this protocol may not have performed as well as the other studies. Despite this constraint, it is still useful for this study to compare this protocol with the other LMM protocols and GMM because it provides a good example of a comprehensive LMM protocol.

My study suggests that GMM, and its statistical toolkit, represents a useful addition to quantitative studies of taxonomy. GMM provides improved insights into the detection of allometric and non-allometric shape variation, usually a confounding factor difficult to untangle with a visual examination or by using LMM. In particular, it is worth commencing a morphological study of a taxon with a GMM protocol that will help to identify those shape patterns that discriminate best. Subsequently, linear morphometric studies can rely on the measurements that account for the relevant shape patterns identified in GMM and be applied to larger sample sizes. Moreover, the overall pattern with GMM and linear morphometrics is that removal of isometry and allometry reduces taxon discrimination and classification performance. Hence, I recommend that further taxonomic endeavours should be more explicit in considering size and allometry for morphological assessments. Otherwise, the risk of splitting allometrically similar taxa can result in artificial group separation.

References

Adams, D. C., Collyer, M. L., Kaliontzopoulou, A., & Sherratt, E. (2016). *geomorph: Software for geometric morphometric analyses*. https://rune.une.edu.au/web/handle/1959.11/21330

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the ‘revolution.’ *Italian Journal of Zoology*, *71*(1), 5–16. https://doi.org/10.1080/11250000409356545

Baker, A. M., & Van Dyck, S. (2013). Taxonomy and redescription of the Yellow-footed Antechinus , *Antechinus flavipes* (Waterhouse) (Marsupialia: Dasyuridae). *Zootaxa*, *3649*(1), 1–62. https://doi.org/10.11646/zootaxa.3649.1.

Bookstein, F. L. (2015). The Relation Between Geometric Morphometrics and Functional Morphology, as Explored by Procrustes Interpretation of Individual Shape Measures Pertinent to Function. *The Anatomical Record*, *298*(1), 314–327. https://doi.org/10.1002/ar.23063

Bookstein, F. L. (2017a). A method of factor analysis for shape coordinates. *American Journal of Physical Anthropology*, *164*(2), 221–245. https://doi.org/10.1002/ajpa.23277

Bookstein, F. L. (2017b). A Newly Noticed Formula Enforces Fundamental Limits on Geometric Morphometric Analyses. *Evolutionary Biology*, *44*(4), 522–541. https://doi.org/10.1007/s11692-017-9424-9

Cardini, A., & Polly, P. D. (2013). Larger mammals have longer faces because of size-related constraints on skull form. *Nature Communications*, *4*, 2458. https://doi.org/10.1038/ncomms3458 https://www.nature.com/articles/ncomms3458#supplementary-information

Claude, J. (2013). Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: Three worked examples in R. *Virtual Morphology and Evolutionary Morphometrics in the New Millenium.*, 94.

Collyer, M. L., & Adams, D. C. (2018). RRPP: An r package for fitting linear models to high‐dimensional data using residual randomization. *Methods in Ecology and Evolution*, *9*(7), 1772–1779.

Dickman, C. R., Parnaby, H. E., Crowther, M. S., & King, D. H. (1998). *Antechinus agilis* (Marsupialia: Dasyuridae), a new species from the *A. stuartii* complex in south-eastern Australia. *Australian Journal of Zoology*, *46*(1), 1–26. https://doi.org/10.1071/zo97036

Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis: With applications in R* (Vol. 995). John Wiley & Sons.

Farkas, L. G., Tompson, B. D., Katic, M. J., & Forrest, C. R. (2002). Differences Between Direct (Anthropometric) and Indirect (Cephalometric) Measurements of the Skull. *Journal of Craniofacial Surgery*, *13*(1), 105–108.

Fruciano, C. (2016). Measurement error in geometric morphometrics. *Development Genes and Evolution*, *226*(3), 139–158. https://doi.org/10.1007/s00427-016-0537-4

Garczarek, U., & Weihs, G. (2003). Standardizing the Comparison of Partitions. *Computational Statistics*, *18*(1), 143–162. https://doi.org/10.1007/s001800300136

Kaliontzopoulou, A., Carretero, M. A., & Llorente, G. A. (2008). Head shape allometry and proximate causes of head sexual dimorphism in *Podarcis* lizards: Joining linear and geometric morphometrics. *Biological Journal of the Linnean Society*, *93*(1), 111–124. https://doi.org/10.1111/j.1095-8312.2007.00921.x

Kendall, D. G. (1989). A survey of the statistical Theory of shape. *Statistical Science*, *4*(2), 87–99. JSTOR.

Klingenberg, C. P. (2016). Size, shape, and form: Concepts of allometry in geometric morphometrics. *Development Genes and Evolution*, *226*(3), 113–137. https://doi.org/10.1007/s00427-016-0539-2

Klingenberg, C. P., Neuenschwander, B. E., & Flury, B. D. (1996). Ontogeny and Individual Variation: Analysis of Patterned Covariance Matrices with Common Principal Components. *Systematic Biology*, *45*(2), 135–150. https://doi.org/10.1093/sysbio/45.2.135

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology*, *36*(2), 235–247. https://doi.org/10.1007/s11692-009-9055-x

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

Outomuro, D., & Johansson, F. (2017). A potential pitfall in studies of biological shape: Does size matter? *Journal of Animal Ecology*, *86*(6), 1447–1457. https://doi.org/10.1111/1365-2656.12732

Pilbeam, D., & Gould, S. J. (1974). Size and Scaling in Human Evolution. *Science*, *186*(4167), 892–901. https://doi.org/10.1126/science.186.4167.892

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*.

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*, *17*(5), 1004–1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Seifert, B. (2008). Removal of allometric variance improves species separation in multi-character discriminant functions when species are strongly allometric and exposes diagnostic characters. *Myrmecological News*, *11*, 91–105.

Sidlauskas, B. L., Mol, J. H., & Vari, R. P. (2011). Dealing with allometry in linear and geometric morphometrics: A taxonomic case study in the *Leporinus cylindriformis* group (Characiformes: Anostomidae) with description of a new species from Suriname. *Zoological Journal of the Linnean Society*, *162*(1), 103–130. https://doi.org/10.1111/j.1096-3642.2010.00677.x

Slice, D. E. (2006). *Modern morphometrics in physical anthropology*. Springer Science & Business Media.

Slice, D. E. (2007). Geometric Morphometrics. *Annual Review of Anthropology*, *36*(1), 261–281. https://doi.org/10.1146/annurev.anthro.34.081804.120613

Stone, J. R. (1997). The spirit of D’Arcy Thompson dwells in empirical morphospace. *Mathematical Biosciences*, *142*(1), 13–30. https://doi.org/10.1016/S0025-5564(96)00186-1

Travouillon, K. J. (2016). Investigating dental variation in *Perameles nasuta* Geoffroy, 1804, with morphological evidence to raise *P. nasuta pallescens* Thomas, 1923 to species rank. *Zootaxa*, *4114*(4), 351–392.

Van Dyck, S., & Crowther, M. S. (2000). Reassessment of northern representatives of the *Antechinus stuartii* complex (Marsupialia: Dasyuridae): *A subtropicus* sp. Nov. And *A. adustus* new status. *Memoirs-Queensland Museum*, *45*(2), 611–635.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer New York. https://doi.org/10.1007/978-0-387-21706-2

Viacava, P., Baker, A. M., Blomberg, S. P., Phillips, M. J., & Weisbecker, V. (2021). Using 3D geometric morphometrics to aid taxonomic and ecological understanding of a recent speciation event within a small Australian marsupial (*Antechinus*: Dasyuridae). *Zoological Journal of the Linnean Society*, *zlab048*. https://doi.org/10.1093/zoolinnean/zlab048

Viacava, P., Blomberg, S. P., Sansalone, G., Phillips, M. J., Guillerme, T., Cameron, S. F., Wilson, R. S., & Weisbecker, V. (2020). Skull shape of a widely distributed, endangered marsupial reveals little evidence of local adaptation between fragmented populations. *Ecology and Evolution*, *10*(18), 9707–9720. https://doi.org/10.1002/ece3.6593

Webster, M., & Sheets, H. D. (2010). A Practical Introduction to Landmark-Based Geometric Morphometrics. *The Paleontological Society Papers*, *16*, 163–188. https://doi.org/10.1017/S1089332600001868

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). KlaR analyzing German business cycles. In *Data analysis and decision support* (pp. 335–343). Springer.

Weisbecker, V., Guillerme, T., Speck, C., Sherratt, E., Abraha, H. M., Sharp, A. C., Terhune, C. E., Collins, S., Johnston, S., & Panagiotopoulou, O. (2019). Individual variation of the masticatory system dominates 3D skull shape in the herbivory-adapted marsupial wombats. *Frontiers in Zoology*, *16*(1), 41. https://doi.org/10.1186/s12983-019-0338-5

Wood, B. A., & Stack, C. G. (1980). Does allometry explain the differences between “Gracile” and “Robust” australopithecines? *American Journal of Physical Anthropology*, *52*(1), 55–62. https://doi.org/10.1002/ajpa.1330520108

Yazdi, A. B. (2014). Application of geometric morphometrics to analyse allometry in two species of the genus *Myrmica* (Hymenoptera: Formicidae). *Soil Organisms*, *86*(1), 77–84.