Integrative and Comparative Biology

Integrative and Comparative Biology, pp. 1–13
doi:10.1093/icb/icz101

Society for Integrative and Comparative Biology

SYMPOSIUM

3D Digitization in Functional Morphology: Where is the Point of Diminishing Returns?

Sharlene E. Santana,[1,*,†] Jessica H. Arbour,[*,†] Abigail A. Curtis[*,†] and Kathryn E. Stanchak[*,†]

*Department of Biology, University of Washington, Seattle, WA 98195, USA; †Burke Museum of Natural History and Culture, University of Washington, Seattle, WA 98195, USA

From the symposium "Comparative Evolutionary Morphology and Biomechanics in the Era of Big Data" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 3–7, 2019 at Tampa, Florida.

[1]E-mail: ssantana@uw.edu

Synopsis Modern computational and imaging methods are revolutionizing the fields of comparative morphology, biomechanics, and ecomorphology. In particular, imaging tools such as X-ray micro computed tomography (μCT) and diffusible iodine-based contrast enhanced CT allow observing and measuring small and/or otherwise inaccessible anatomical structures, and creating highly accurate three-dimensional (3D) renditions that can be used in biomechanical modeling and tests of functional or evolutionary hypotheses. But, do the larger datasets generated through 3D digitization always confer greater power to uncover functional or evolutionary patterns, when compared with more traditional methodologies? And, if so, why? Here, we contrast the advantages and challenges of using data generated via (3D) CT methods versus more traditional (2D) approaches in the study of skull macroevolution and feeding functional morphology in bats. First, we test for the effect of dimensionality and landmark number on inferences of adaptive shifts during cranial evolution by contrasting results from 3D versus 2D geometric morphometric datasets of bat crania. We find sharp differences between results generated from the 3D versus some of the 2D datasets (*xy*, *yz*, ventral, and frontal), which appear to be primarily driven by the loss of critical dimensions of morphological variation rather than number of landmarks. Second, we examine differences in accuracy and precision among 2D and 3D predictive models of bite force by comparing three skull lever models that differ in the sources of skull and muscle anatomical data. We find that a 3D model that relies on skull μCT scans and muscle data partly derived from diceCT is slightly more accurate than models based on skull photographs or skull μCT and muscle data fully derived from dissections. However, the benefit of using the diceCT-informed model is modest given the effort it currently takes to virtually dissect muscles from CT scans. By contrasting traditional and modern tools, we illustrate when and why 3D datasets may be preferable over 2D data, and vice versa, and how different methodologies can complement each other in comparative analyses of morphological function and evolution.

Introduction

The fields of comparative functional morphology and biomechanics are being revolutionized by greater access to and the development of imaging technologies and methods to explore morphology *in silico* and in three dimensions (e.g., micro computed tomography, μCT; diffusible iodine-based contrast enhanced μCT, diceCT; high-resolution laser scanning). By using these tools, researchers have produced some of the most detailed and massive morphological datasets for vertebrates to date, which have already proven critical to understanding the function and diversity of many key morphological structures (e.g., Sykes et al. 2016; Bardua et al. 2019; Felice et al. 2019). Simultaneously, computational advances have enabled high throughput of large morphological datasets produced via more traditional methods to quantify morphology, including linear measurements, photography, and histology (e.g., Montgomery et al. 2011; Gehan et al. 2017). In addition, increasing availability of data-rich phylogenetic hypotheses allows more robust tests of functional questions within an evolutionary context.

Via their impact on functional morphology research, imaging and computational resources are advancing the field of ecomorphology in new and exciting directions. Ecomorphology deals with investigating how variation in morphology and behavior results in variation in performance (the ability to conduct fitness-relevant tasks), and how differences in performance enable niche partitioning (Arnold 1983). In the case of dietary ecomorphology, such a research program generally involves conducting comparative analyses of skull and/or jaw muscle morphology, documenting feeding behavior and/or kinematics, and modeling or measuring feeding performance metrics (e.g., bite force (BF), gape, and suction speed) (Wainwright 1994; Irschick and Higham 2016). For small vertebrates, μCT scanning has produced unprecedented results in the quantification of external and internal skull morphology (e.g., Phillips et al. 2009; Curtis and Simmons 2017), documentation of jaw muscle architecture (e.g., Jeffery et al. 2011; Dickinson et al. 2018; Ford et al. 2018; Santana 2018), and the construction of biomechanical models to predict feeding performance (Davis et al. 2010; Cox et al. 2011; Peterson and Müller 2018), all of which has contributed to a better understanding of the dietary ecomorphology and diversification of some of the most ecologically diverse vertebrate lineages.

Despite these advances, it is still unclear when modern three-dimensional (3D) methods should be chosen over more traditional approaches (e.g., photography-based skull shape analyses, linear measurements) to quantify and compare cranial morphology, function, and evolution. Given the costs associated with μCT data acquisition and processing, this could prove important for planning cost-effective functional morphology research. As part of the symposium "Comparative Evolutionary Morphology and Biomechanics in the Era of Big Data," the goal of this paper is to showcase whether and when modern imaging tools, and the large datasets they generate, may be preferred when trying to uncover patterns of morphological evolution and model anatomical function. To do so, we use bats as a study system in two case studies that illustrate workflows in modern ecomorphological research. First, we contrast results from macroevolutionary analyses of 2D and 3D skull shape datasets to explore how data dimensionality and specimen orientation can affect the estimation of evolutionary patterns. Second, we compare predictive models of bite performance that integrate various levels of traditional and modern sources of morphological data. Bats are a morphologically and ecologically diverse group of small mammals, and are therefore an excellent natural experiment within which to test the sensitivity of different tools to morphological variation.

Mapping skull shape evolution across bats: 3D versus 2D geometric morphometrics

Among mammals, bats exhibit extraordinary cranial diversity. Some of this diversity may be adaptive and thus directly related to function, performance, and ecology, and some may be the result of non-adaptive processes. These sources of variation can be teased apart through phylogenetic comparative analyses that test for selective regime shifts in skull shape, as these can reveal evolutionary patterns in morphology that may be the result of ecological adaptation (Butler and King 2004). To inform these analyses, skull shape can be quantified through geometric morphometric (landmark-based) methods that produce variables describing size-independent shape differences (Zelditch et al. 2004). Data collection for these methods can be done in 2D from photographs or radiographs of different views of the skull, or in 3D from reconstructions based on μCT or laser scans of skulls (but see Olsen and Westneat 2015). The advantage of either approach is an active subject of study (e.g., Cardini 2014; Openshaw et al. 2017; Buser et al. 2018).

To contrast the effects of using morphological data derived from 2D versus 3D approaches in macroevolutionary analyses, we used an existing geometric morphometric dataset of bat skulls that spans 202 species ($N = 1$–8 per species), 50–100% generic coverage within bat families, ~20% total species coverage, and all diets and sensory modalities within Chiroptera (Arbour et al. 2019). To build this dataset, we (1) μCT scanned skulls from museum specimens using a Skyscan 1172 (Skyscan, Belgium), (2) reconstructed μCT slices using NRecon (Microphotonics Inc., Allentown, PA), (3) segmented bone tissue and exported *.stl (surface files) of the skulls using Mimics (Materialize, Ann Arbor, MI), (4) cleaned and reduced mesh size in Geomagic Studio (3D Systems, Rock Hill, SC), and (5) placed homologous landmarks and sliding semi-landmarks on the skull surface using Checkpoint v.2017 (Stratovan, Davis, CA) (Fig. 1). More details about this dataset can be found in Arbour et al. (2019). Focusing on the cranium, and to compare results of evolutionary analyses on 3D versus 2D geometric morphometric datasets, we used the full, 3D dataset in one set of analyses, and reduced ("2D") versions of this dataset in several additional analyses. We created two versions of the 2D datasets to examine the

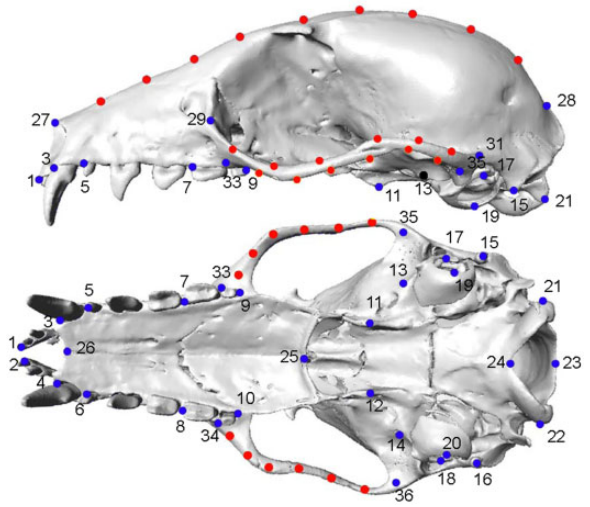

Fig. 1 Location of 3D landmarks (numbered) and sliding semilandmarks used in geometric morphometric analyses of the bat cranium.

relative impacts of the loss of dimensionality and the reduction of landmarks on the results of macroevolutionary analyses. First, to test the effect of reduced landmark dimensionality alone, we created three 2D datasets by removing either the *x*, *y*, or *z* coordinate components of the landmarks and semilandmarks—as the initial landmark configurations were aligned with the major anatomical axes. These datasets would mimic 2D data collection from frontal, ventral, or lateral radiographs of the cranium, respectively.

Second, to test the additional effect of changes in the number of landmarks that can be placed on each 2D view of the cranium (e.g., when photographs are used for data collection), we created three additional datasets by removing the landmarks and semilandmarks that would not be visible in photos of the frontal, ventral, and lateral views of the cranium (herein referred to as frontal, ventral, and lateral views). The landmarks included for each of these views are detailed in Table 1. Bat skulls vary substantially in shape and not all of these landmarks may be visible in all specimens, so we selected landmarks that would be visible in at least 10% of all species. Using the same original dataset to derive reduced datasets, as opposed to creating new ones from radiographs or photographs of crania with 2D digitization, allowed us to visualize the effect of loss of dimensionality or landmarks without introducing the additional error that is inherent to digitizing landmarks in two versus three dimensions (Robinson and Terhune 2017).

We ran a generalized Procrustes analysis on each of the seven landmark datasets (3D, *xy*, *xz*, *yz*, and simulated frontal, ventral, and lateral views) to remove the effects of scale, rotation, and position (Zelditch et al. 2004). Aligned landmark configurations were averaged by species and across bilaterally symmetrical landmarks after mirroring. We used phylogenetic principal component analysis (pPCA; Revell 2012) to examine the major patterns of variation across the aligned landmark datasets while accounting for evolutionary relatedness, using the R function *phyl.pca*. We used parallel analysis (Horn 1965) as a stopping rule to select a number of axes for consideration, following the implementation by Peres-Neto et al. (2005) (e.g., using the 95th percentile of simulated values; see Supplementary Material). We calculated cutoff values per dataset using the relative eigenvalues from pPCA of 100 simulated datasets of random, uncorrelated variables with variances equal to the observed shape data. We used a Brownian motion model on all pPCAs due to computational limitations for re-scaling branch lengths in *phyl.pca*. We used the pPCA scores generated for each of the 3D and 2D datasets in subsequent shape analyses.

To test the impact of removing non-visible landmarks on shape results, we used a two-block partial least squares test (R package geomorph, function *two.b.pls*; Adams and Otárola-Castillo 2013) comparing each of the three pairs of landmark datasets matched by view (i.e., *xy* vs. ventral, *xz* vs. lateral, and *yz* vs. frontal). We also examined whether each of the resulting pPCA axes showed similar patterns of morphological variation across species by regressing the scores of each phylogenetic principal component (pPC) axis from the ventral, lateral, and frontal datasets on the *xy*, *xz*, and *yz* datasets, respectively. Lastly, we contrasted the results of the 3D and 2D datasets, both on the original landmarks and pPCA scores in two ways: (1) a two-block partial least squares (PLS) test to contrast the 3D landmarks with each of the six 2D landmark datasets (R function *two.b.pls* in geomorph), and (2) a Mantel test on Euclidean distance matrices from the critical pPC scores (function *mantel.rtest*, ade4 package).

To investigate if there are differences in the estimated patterns of cranial evolution when different datasets are used, we estimated the configuration of adaptive shifts through the l1ou method (Khabbazian et al. 2016). This method uses the Ornstein–Uhlenbeck (OU) process to model a changing adaptive landscape over time and over lineages, and selects evolutionary shifts using a "lasso" approach and no a priori assumptions of the number or location of adaptive shifts (Khabbazian et al. 2016). We

Table 1 Landmarks and semilandmarks included in each of the simulated lateral, ventral, and frontal view datasets

	Lateral	Ventral	Frontal
Landmarks	1:10, 13:22, 27:36	1:26	1:7, 26, 27, 29, 30, 33, 34
Semilandmark Series	Zygomatic dorsal (all), zygomatic ventral (all), dorsal midline (all)	Zygomatic ventral (all)	Zygomatic dorsal (anteriormost 5), zygomatic ventral (anteriormost 4), dorsal midline (anteriormost 7)

See Fig. 1 for landmark locations.

performed these analyses on the pPC scores of all critical axes for each of the multivariate morphospaces (below) in R v.3.3.3 using functions in the package geomorph (Adams and Otárola-Castillo 2013), l1ou (Khabbazian et al. 2016), and phytools (Revell 2012). We used the l1ou method and pPC scores for these analyses because, to date, there are no established comparative methods to conduct analyses of adaptive regime shifts (using OU models) on high-dimensional shape data. Since it is currently not computationally feasible to carry out such adaptive shift fitting analyses on the full geometric morphometric landmark data, some reduction of trait dimensionality—through PCA or similar approaches—is necessary. It is important to note that all currently available approaches for automatic detection of adaptive shifts for multivariate data (l1ou, surface, phylogenetic EM, etc.) implement a reduced version of the multivariate OU model, excluding either trait covariation or assuming that selective constraints are constant across traits. In addition, as shown by Adams and Collyer (2018), all methods exhibit exceptionally high rates of model misspecification, and analyses limited to the first few PCs/pPCs tend to be biased. In the case of pPCA, this bias depends on the fit of the underlying Brownian model to the data (Uyeda et al. 2015). Moreover, analyses that remove phylogenetic correlations by pPCA before using a method designed for independent traits (such as l1ou) can be misleading in the presence of shifts (Bastide et al. 2018). Here, we emphasize the differences in macroevolutionary results inferred by differences in the dimensionality of landmark coordinates, rather than other specific conclusions. Nevertheless, we doubt that the differences between 2D and 3D coordinate data in our results are merely methodological artifacts.

Our results (Figs. 2 and 3) illustrate that the morphological and macroevolutionary patterns estimated by l1ou are highly dependent not only on the dimensionality of the geometric morphometric dataset (2D versus 3D), but which views (frontal, ventral, or lateral) are used to capture cranial morphology in 2D. We found strong relationships between landmark coordinate datasets with and without visible landmarks (two-block PLS: all three $r > 0.99$, all three $P = 0.001$). Similarly, each of the matched 2D datasets (i.e., those with and without visible landmarks in the same anatomical plane) showed similar major axes of morphological variation from pPCA. Scores from each of these respective axes were strongly correlated (Table 2), reflecting the shape changes in the anatomical structures favored by each view. The *yz* and simulated frontal datasets recover only one critical axis, describing primarily rostral flexure. In contrast, the *xy* and simulated ventral datasets recover two critical axes, which describe rostral elongation (pPC1) and width of the zygomatic and basicranium (pPC2). The *xz* and simulated lateral datasets also recover two axes, which describe the degree of rostral elongation (pPC1) and flexure (pPC2) across bats. Logically, the 3D dataset is able to capture all the major patterns of morphological diversity contained in the six 2D datasets, which are summarized in three major axes of variation (Table 2). These describe the degree of rostral elongation (pPC1) and flexure (pPC2), and skull height (pPC3). Qualitatively, many of the 2D pPC axes overlap with major morphological patterns seen in the 3D analyses (e.g., rostral elongation and flexure; Fig. 3). However, the 3D analysis further captures an axis of variation (skull height) not observed among the major axes recovered by the 2D analyses. Quantitative comparisons of both the landmark coordinate data and the pPCA scores further demonstrated significant relationships between the 2D and 3D datasets, with each of the 2D datasets varying in its degree of correlation with the full 3D dataset. Both the two-block PLS and the Mantel tests indicated that the lateral/*yz* configurations have the strongest correlation with the 3D data, followed by the frontal/*xz* datasets, and then the ventral/*xy* dataset (Table 3).

The results from the l1ou analyses further illustrate that using only the *xy*, *yz*, frontal, or ventral view datasets leads to recovering patterns of adaptive evolution that are sharply different from those resulting from analyses of the 3D dataset (Figs. 2

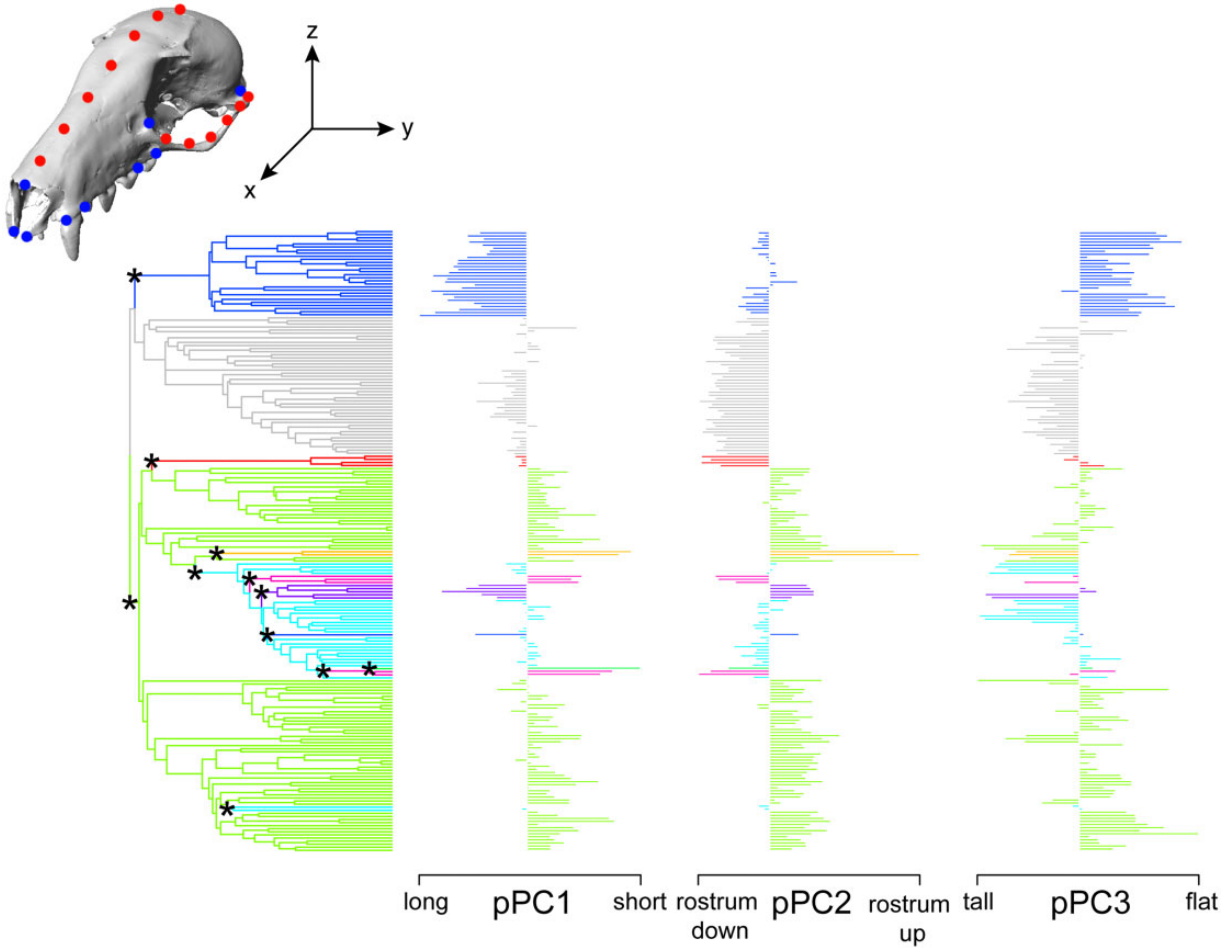

Fig. 2 Evolutionary shifts (*) in cranium shape across bats, as determined by l1ou adaptive landscape model fitting on pPCA scores from the full 3D dataset. Morphological variation along critical (pPC) axes is illustrated by the barplots to the right of the phylogeny.

and 3A, B, E, F). However, the same analyses on the *xz* and lateral view datasets recovered nearly identical results to those from the 3D dataset (Figs. 2 and 3C, D). This finding makes intuitive sense if we consider that the major traits that underlie morphological patterns in 3D (rostral elongation and flexure) are also captured by the *xz* and lateral view data. Conversely, as these traits are not completely captured by the frontal and ventral views (or *xy*, *yz* datasets), these yield dramatically different interpretations about the adaptive evolution of cranial morphology in bats. Evolutionary analyses on the ventral and frontal views showed few differences with the *xy* and *yz* dataset, respectively (Fig. 3). Thus, the differences observed between the 3D dataset versus some of the 2D datasets (*xy*, *yz*, ventral, and frontal) do not appear to be primarily driven by the decrease in the number of landmarks used to generate each dataset, but rather by the loss of critical dimensions of morphological variation. This suggests that 2D radiographs in some views are unlikely to capture sufficient shape variation when compared with 3D approaches.

Ventral and lateral views of the cranium are the gold standard in 2D geometric morphometric analyses of mammal skulls (e.g., Figueirido et al. 2013; Xia et al. 2013; Linde-Medina et al. 2016). Our results highlight that, for clades that have similar morphological trends as bats, the ventral view by itself is not sufficient to estimate patterns of cranial evolution and making inferences about the potential ecological forces shaping cranial diversity. Similar issues have been documented by other studies of cranial morphology in which the shape information missing from the 2D dataset had a strong impact on statistical tests of ecomorphological hypotheses (Buser et al. 2018). Since the results from the lateral view are consistent with those from the 3D dataset, a well-planned 2D geometric morphometric study could potentially overcome some of these issues. However, the selection of appropriate views for 2D analyses might not be straightforward in cases where

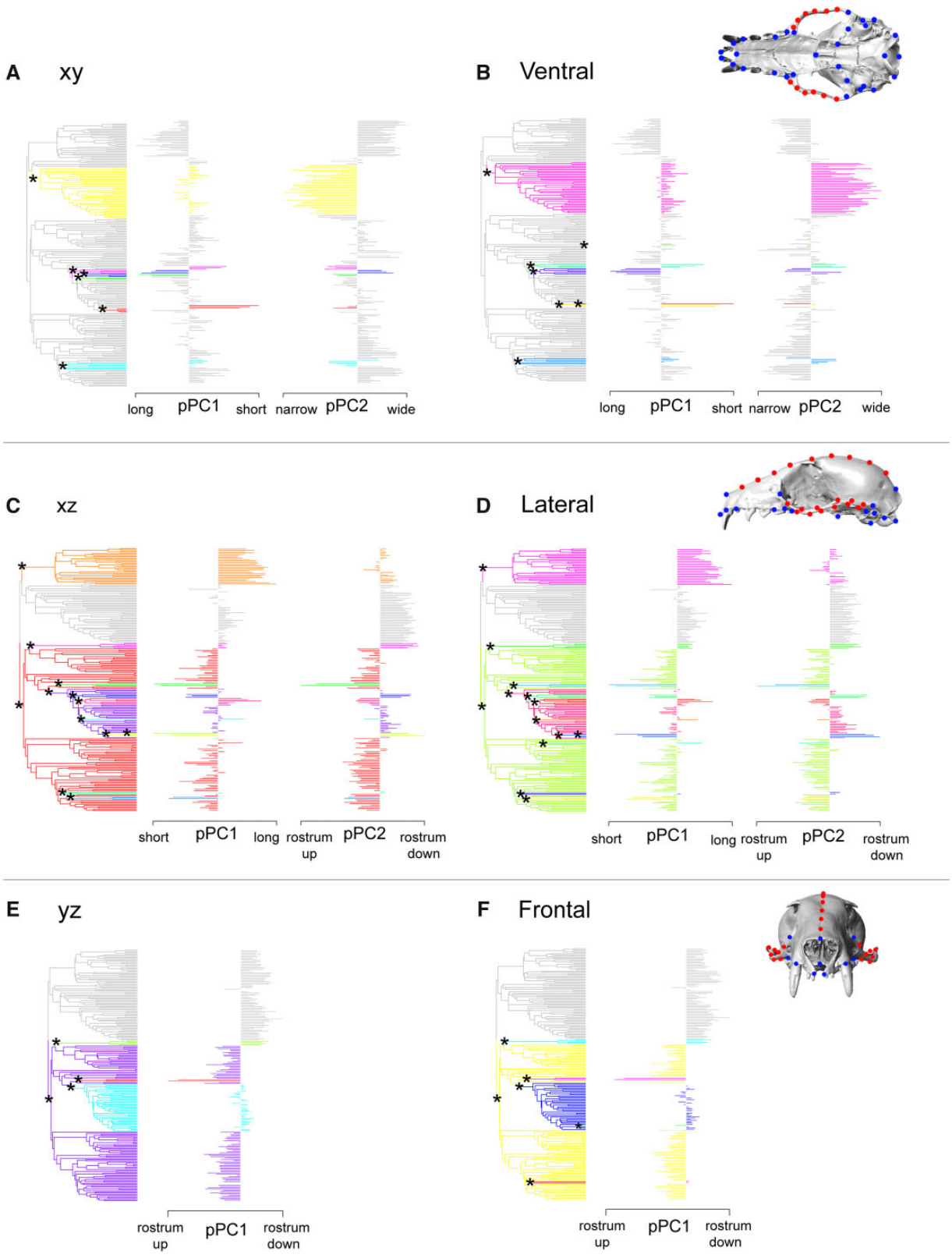

Fig. 3 Comparisons of evolutionary shifts (*) in cranium shape across bats, as determined by l1ou adaptive landscape model fitting on pPCA scores from the 2D and visible landmarks datasets (see text): (**A**) *xy*, (**B**) simulated ventral view, (**C**) *xz*, (**D**) simulated lateral view, (**E**) *yz*, and (**F**) simulated frontal view. Morphological variation along critical (pPC) axes is illustrated by the barplots to the right of each phylogeny.

Table 2 Summary of pPCA derived from geometric morphometric analyses on the 2D and 3D datasets

Dataset	Percent variation			Dataset	Percent variation			Correlation (*r*)	
	pPC1	pPC2	pPC3		pPC1	pPC2	pPC3	pPC1	pPC2
3D	37.4	22.5	9.9						
xy	33.9	31.4	–	**Ventral**	37.5	27.4	–	0.971	0.976
xz	41.2	25.1	–	**Lateral**	49.6	18.2	–	0.986	0.976
yz	59.5	–	–	**Frontal**	63.0	–	–	0.992	NA

Percent variation, percentage of variation explained by each of the critical pPC axes; Correlation, correlation coefficients from linear regression between pPC scores from each of the matched datasets with and without visible landmarks (e.g., *xy* vs. ventral). All regressions were significant after a Bonferroni correction of α values ($P < 0.01$).

Table 3 Comparisons of the 3D and each of the 2D landmark configurations

	r	*P*		*r*	*P*
Two-block PLS					
xy	0.919	0.001	**Ventral**	0.937	0.001
xz	0.998	0.001	**Lateral**	0.997	0.001
yz	0.974	0.001	**Frontal**	0.967	0.001
Mantel test					
xy	0.8379	0.001	**Ventral**	0.7951	0.001
xz	0.9805	0.001	**Lateral**	0.9669	0.001
yz	0.8533	0.001	**Frontal**	0.8484	0.001

Two-block PLS tests were carried out on the landmark coordinates. Mantel tests were carried out on Euclidean distance matrices calculated from the pPC scores (see Table 2). *r*, correlation coefficient; *P*, significance from randomization tests.

clade-wide trends in morphological variation of a complex structure are unknown. Therefore, we recommend using 3D methods in comparative analyses of bony structures due to their ability to capture morphological variation comprehensively and without the need for a priori assumptions about major trends. Although some forms of 3D data acquisition (e.g., µCT scanning) are still costly, methods such as stereo-photography provide a practical, low-cost, and fast approach to 3D morphometric analyses (e.g., StereoMorph; Olsen and Westneat 2015), and repositories of 3D data (e.g., Morphosource) are rapidly expanding due to increasing digitization efforts.

Estimating bite performance in bats: traditional versus digital methods

In recent years, 3D biomechanical modeling has allowed ecomorphologists to fine-tune the links between dietary demands and morphological variation in the feeding apparatus across a wide range of vertebrates (e.g., Slater et al. 2009; Dumont et al. 2011; Watson et al. 2014; Santana 2016). Methodologically, constructing these models lies at the interface of modern and traditional comparative morphology; they integrate 3D data on skull shape (e.g., from µCT scans) with data on jaw muscle anatomy (e.g., mass, attachment sites, fiber lengths and orientations). The latter are primarily derived from dissections, which—as a destructive sampling approach—limits the number of specimens and species that can be incorporated into these analyses. In recent years, however, methods such as diceCT (Gignac et al. 2016) have facilitated the non-destructive documentation of jaw muscle anatomy *in situ*, in 3D, and in great detail (e.g., Cox and Faulkes 2014; Gignac and Kley 2014; Dickinson et al. 2018; Santana 2018), thereby holding great promise to study rare specimens and/or generate high-resolution, 3D anatomical datasets for potentially more accurate biomechanical models (e.g., Cox et al. 2011).

Currently, data generation via diceCT involves a substantial cost, primarily because segmentation of muscle tissue from CT scan slices still has to be performed manually. For example, it can take 10–30 h to digitally segment the jaw musculature of a very small (~2cm) bat head that has been scanned at a high enough resolution to document skull detail (~10–30 µm). This is substantially greater than the time it would take to collect the same data (muscle mass and attachment areas) through a dissection (0.5–1 h, on average). Second, the staining time to achieve appropriate tissue contrast is still not easily predictable, and seems to depend on specimen size, age, and preservation method (Gignac and Kley 2014, 2019; Santana 2018). Third, although there have been advances in the automated detection and measurement of muscle fiber lengths from diceCT scans (Jeffery et al. 2011; Dickinson et al. 2018), these methods still need refinement, and thus muscle fiber length data still need to be generated via dissection. Finally, jaw adductor physiological cross-sectional areas (PCSA) partly derived from diceCT data are equivalent to those generated fully from dissection (Santana 2018). All these issues raise the

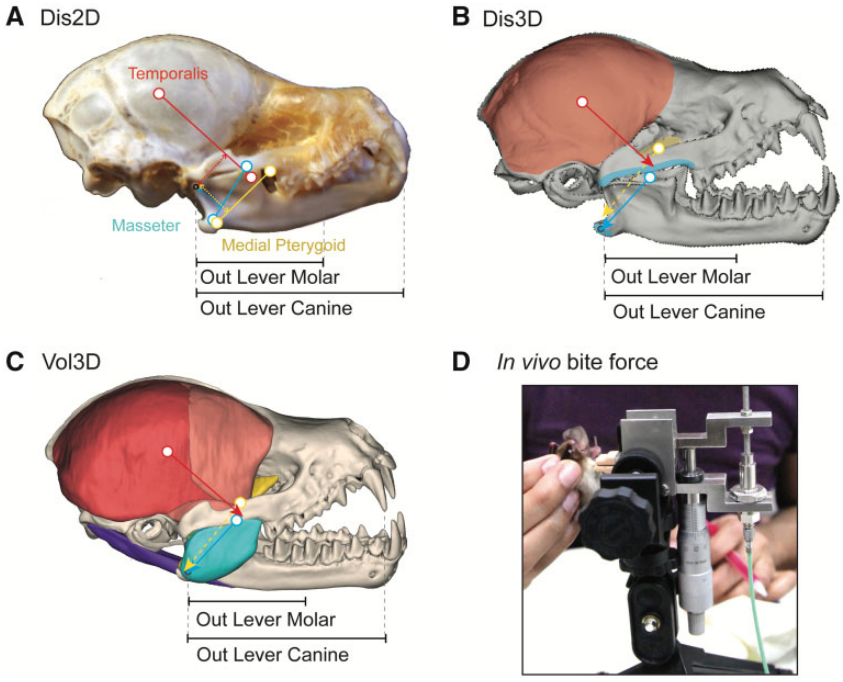

Fig. 4 Illustration of the models used to estimate BF, using the skull of the bat *Trachops cirrhosus* as an example: (**A**) Dis2D method, with jaw adductor lines of action (solid lines), moment arms (dashed lines), and out levers calculated from a skull photograph; (**B**) Dis3D method, with jaw adductor attachment regions defined on a 3D model of the cranium, and mandible and muscle forces (arrows) directed toward the respective insertion area centroids on the mandible; (**C**) Vol3D method, a modification of the Dis3D method in which muscle forces are calculated from muscles volumes derived from diceCT scans. BF predictions from these models were compared with *in vivo* BFs from wild bats, measured using a piezoelectric transducer setup (**D**).

question: when are the costs of using modern imaging tools outweighed by their benefits when building predictive models of feeding performance?

To identify the advantages of using diceCT versus exclusively traditional methods in informing predictive models of bite performance in bats, we compared the accuracy and precision of three models that differed in the sources of skull and muscle anatomical data (Fig. 4). These models rely on the principle that the jaw of most mammals approximates a third-class lever system (Hylander 1975), in which the in levers are defined by attachments of the jaw adductors (temporalis, pars suprazygomatica, zygomaticomandibularis, masseter, medial and lateral pterygoids), the fulcrum is the temporomandibular joint (TMJ), the out lever is defined by the biting point along the mandible, and forces applied are proportional to the PCSAs of the jaw adductors. We applied these models across 10 species of morphologically distinct noctilionoid bats (Table 4), and sourced all input data (μCT skull data, dissection PCSAs, and diceCT PCSAs) from Santana (2018).

Dis2D model: dissection PCSA and 2D skull method

We estimated BFs across species by applying a modification of a method that combines linear measurements from images of the skull with dissection data (Hartstone-Rose et al. 2012) using the equation:

$$\mathrm{BF} = 2 \times \text{Muscle Stress} \times \left(\frac{(\mathrm{PCSAm} \times \mathrm{MAm}) + (\mathrm{PCSAt} \times \mathrm{MAt}) + (\mathrm{PCSAp} \times \mathrm{MAp})}{\text{Out Lever}} \right)$$

where Muscle Stress is 25 N/cm^2, the median value for mammalian muscle at body temperature (Herzog 1994); PCSAm, PCSAt, and PCSAp are the PCSA for the masseter, temporalis, and medial pterygoid, respectively (from Santana 2018); MAm, MAt, and MAp are the moment arms for the masseter, temporalis, and medial pterygoid, respectively, which were measured as the perpendicular distance from the TMJ to each muscle's line of action (Fig. 4A); and Out Lever is the distance between the TMJ and the tip of the canine for canine bites, and the center of the first molar for molar bites. We used ImageJ (Schneider et al. 2012) to measure moment arms and out levers from lateral skull photos. Our modified equation multiplies the BF calculation by 2 to account for bilateral symmetry, since we could not calculate forces generated on the balancing side (as in Hartstone-Rose et al. 2012) due to the lack of electromyography data for the species studied.

Table 4 Canine and molar BFs (in Newtons) estimated by the three methods applied in this study

	Canine				Molar			
Species	**Measured**	**Dis2D**	**Dis3D**	**Vol3D**	**Measured**	**Dis2D**	**Dis3D**	**Vol3D**
Artibeus lituratus	16.343 ± 7.212	7.285	6.147	6.843	21.754 ± 8.532	10.55	10.012	11.145
Artibeus phaeotis	2.837 ± 1.437	1.352	1.252	1.666	5.631 ± 1.630	2.638	1.871	2.490
Carollia perspicillata	5.755 ± 2.461	2.338	2.051	2.850	7.929 ± 2.711	3.442	3.079	4.280
Desmodus rotundus	2.901 ± 1.500	2.587	2.646	3.332	6.063 ± 0.368	4.437	3.545	4.464
Glossophaga soricina	1.286 ± 0.914	0.391	0.471	0.362	1.366 ± 0.700	0.711	0.746	0.574
Micronycteris hirsuta	7.468 ± 1.924	3.926	3.072	4.109	12.480 ± 3.604	6.184	5.145	6.880
Noctilio leporinus	19.900 ± 8.910	7.000	7.124	9.113	–	11.093	9.813	12.552
Pteronotus parnellii	3.890 ± 2.347	6.229	2.678	2.969	7.758 ± 0.999	8.760	4.149	4.600
Sturnira lilium	5.724 ± 2.792	1.266	1.697	1.814	8.055 ± 3.737	2.508	2.428	2.596
Trachops cirrhosus	8.009 ± 2.611	4.649	3.522	4.307	11.741 ± 4.667	8.316	5.401	6.604

Averages of measured maximum BFs (from Aguirre et al. 2002; Santana et al. 2010; Santana 2016) are shown for comparison. Dash indicates no BF value available.

Dis3D: dissection PCSA and 3D skull method

We estimated BFs by applying a custom-written R program (BiteR; Supplementary Material) that combines data on forces of all jaw adductor muscles, generated from dissections, with a 3D representation of the skull generated through μCT. As input, the model requires (1) centroids of the muscle attachment areas on one side of the skull and the two TMJ fossae, (2) PCSAs of each of the adductor muscles, (3) the side of the skull of the considered muscles, (4) a value for muscle stress (in this case, 25 N/cm^2), and (5) an out lever measurement. The centroids of the TMJ fossae and the side of the skull are used to define a TMJ axis, whereas the other data are used to define the torques generated by the muscles about the TMJ axis and the total BF. This model is equivalent to that used by Davis et al. (2010), with the exception that BiteR considers the entire muscle force to be applied at the centroid of the muscle attachment on the cranium and does not distribute the applied muscle force across the entire cranial attachment area. The full code with detailed input data is provided in the Supplementary Material.

Based on photo documentation during dissections, we defined the TMJ fossae and muscle origin and insertion areas on surface files (*.stl) of the cranium and mandible, which were derived from μCT scans (Santana 2018). The surface files of the cranium and mandible had first been set at a 30° gape to match the *in vivo* measurements taken with a BF meter (Fig. 4D). We exported the attachment areas as binary 3D surface (*.stl) files (Fig. 4B) and then used the *readstl* function (rgl R package) as part of a custom function to calculate the centroids of the stls by geometric decomposition of the triangles defined by the stl file. The full code for calculating these centroids is provided in the Supplementary Material.

BiteR outputs moments about the TMJ axis by each jaw adductor and a total BF estimate that is the sum of all torques produced by the muscles, divided by the defined out lever, and multiplied by 2 to account for bilateral symmetry. We measured out levers from 3D skull models in Geomagic (3D Systems, Rock Hill, SC) as the distance from the TMJ to either the tip of the canine or the center of the occlusal surface of the first molar (Fig. 4B).

Vol3D: diceCT PCSA and 3D skull method

We applied BiteR as in the Dis3D method, but with jaw adductor attachment area and muscle volume data derived from diceCT scans (from Santana 2018; Fig. 4C). To do so, we defined each jaw adductor origin and insertion areas at the regions of contact between the meshes of each muscle and the cranium and mandible. We calculated the volume of each muscle mesh in Geomagic Studio (3D Systems, Rock Hill, SC), and used these data to calculate muscle forces (PCSA = muscle volume/fiber length; Muscle Force = PCSA × 25 N/cm^2). We ran BiteR with these data as described above.

For most species in our dataset, we compiled data on mean *in vivo* maximum canine and molar BFs from our previous published work (Santana et al. 2010; Santana 2016). In these studies, we used a piezoelectric force transducer to measure BFs from free-ranging animals at a 30° gape angle (Fig. 4D). For *Noctilio leporinus*, we used maximum canine BF values from Aguirre et al. (2002), who used the same BF meter setup. To compare the accuracy of the BF

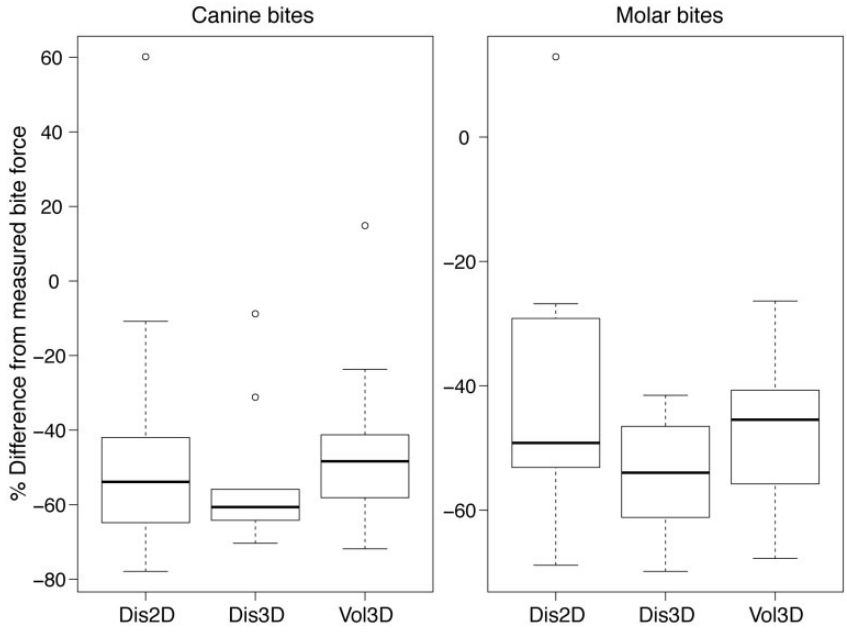

Fig. 5 Error (difference from measured BF as a percentage of measured BF) across methods (from Fig. 4A–C) and bite types tested in this study ($n = 10$ species per method). Error bars represent the upper and lower quartiles.

predictive methods (Dis2D, Dis3D, and Vol3D), we used paired *t*-tests to explore if there was a significant difference between the model predictions and *in vivo* BFs, and a two-way ANOVA and Tukey's honest significant difference test to determine if there was a significant difference in BF prediction error among the three models and across species. We calculated the BF prediction error as a percentage of mean *in vivo* maximum BFs as follows: Error = ([estimated BF–measured BF]/measured BF) × 100). All statistical analyses were conducted in R (R Development Core Team 2019).

All predictive methods significantly underestimated *in vivo* BFs ($P < 0.05$ in all paired *t*-tests; Fig. 5). On average, Vol3D underestimated *in vivo* BFs the least, followed by Dis2D and Dis3D. The three methods did not differ significantly in their error in predicting canine BFs (ANOVA: $P = 0.313$), but they did differ in their error in predicting molar BFs (ANOVA: df = 2, SS = 894, MS = 446.8, $F = 3.596$, $P = 0.04854$). Dis2D had significantly lower error than Dis3D for molar BFs (difference: 13.28%; $P = 0.0399$), although this was driven by one outlier (Fig. 5). Dis3D and Vol3D did not differ significantly from each other in their BF error ($P > 0.05$ in pairwise comparisons for canine and molar bites). Across all methods, there was a statistically significant effect of skull morphology (species) in error magnitude (Canine bites: $P = 0.0003$; Molar bites: $P = 0.006$). Our sample size did not allow us to test for the interaction between method and species. In all, these results suggest that the three methods predict BF with a similar degree of error, although the Dis3D method appears to do so with slightly greater precision. That is, this method was relatively more consistent in the degree of BF underestimation across different morphologies than the other two methods.

Although the 3D methods did not differ statistically in their amount of predictive error, Vol3D provided higher BF estimations than Dis3D. This is because muscle volumes (from diceCT scans) result in greater PCSA calculations (Santana 2018), and thus muscle force estimates, in Vol3D; Dis3D relies on muscle mass and a muscle density constant from the literature for PCSA calculations. The higher accuracy provided by Vol3D is modest, however, especially considering that building these models requires considerably greater effort (i.e., manually segmenting muscles from diceCT scans is a time-consuming task). Nevertheless, diceCT scans can enable researchers to explore aspects of muscle anatomy (e.g., fiber density and architecture, Jeffery et al. 2011; Dickinson et al. 2018), which are difficult to examine with dissections alone. These data could eventually improve the predictions of BF models. Importantly, while the 3D methods modeled BF at the same gape at which it was measured from live animals, Dis2D calculated BF at dental occlusion; this allows controlling for changes in muscle stretching that cannot be easily accommodated by the model (Hartstone-Rose et al. 2012). BF decreases as

gape increases in mammals (Herring and Herring 1974; Dumont and Herrel 2003; Williams et al. 2009; Santana 2016); thus, some of the relatively higher BF estimates from Dis2D are partially an artifact of the modeled gape. In turn, the lower precision of this method may be partly a product of not including some of the smaller jaw muscles and their forces (e.g., pterygoids, zygomaticomandibularis), as their attachment areas are difficult to define in photographs. Jaw muscles vary significantly in their relative importance for BF production across bat species and diets (Herrel et al. 2008; Santana et al. 2010). Therefore, Dis2D may be less desirable when comparing BF predictions across species that differ dramatically in cranial morphology. However, this approach could still be adequate for datasets composed of more morphologically similar species.

The factors that cause the 3D models to have low accuracy in BF estimation are much less clear. Since there is a linear relationship between muscle stress and BF predictions in these types of 3D lever models (Davis et al. 2010), our results suggest that a higher stress value could be used to scale jaw adductor PCSA to forces. However, increasing muscle stress to the highest values reported for mammals (27.5 N/cm^2; Close 1972) does not account for the large discrepancies between the *in vivo* and modeled data. In a previous study, Santana et al. (2010) used the same individual bats to collect *in vivo* BFs and generate anatomical data for a 3D modeling approach akin to BiteR. In that study, we found that the model also underestimated *in vivo* BFs—albeit to a lesser degree—but regressions between model predictions and measurements had a slope not significantly different from 1. Thus, some of the error in Dis3D and Vol3D predictions could be attributed to intraspecific morphological variation, since we did not use the same individuals for *in vivo* measurements. Importantly, simple lever models do not fully account for the complex arrangement of mammalian jaw adductors and their internal architecture, including the effect of muscle wrapping and stacking, and fiber orientation. Since testing those factors is beyond the scope of this paper, future studies could build upon models like BiteR to evaluate the effect of braincase curvature and focal node location on BF estimates, and the value of incorporating fiber orientation data to calculate muscle vectors (e.g., Watson et al. 2014).

Altogether, our findings from this modeling case study emphasize several advantages and shortcomings of applying 2D and 3D modeling approaches in functional morphology and biomechanics research. Although μCT and diceCT provide outstanding levels of anatomical detail, there are still significant costs, challenges, and a need for refinement of models to effectively translate these massive morphological datasets into accurate predictions of performance metrics.

Acknowledgments

The authors thank the staff of the AMNH, FMNH, and the Burke Museum for facilitating access to specimens, and Abby Vander Linden for assistance with dissections and digital segmentation of specimens. Finally, the authors are grateful to Drs Martha Muñoz and Samantha Price for the invitation to present at the "Comparative Evolutionary Morphology and Biomechanics in the Era of Big Data" symposium, which was funded by an NSF Rules of Life FELS grant (IOS 1839250) to the symposium organizers.

Funding

This work was funded by NSF award #1557125 to S.E.S., an AMNH Gerstner Scholars Postdoctoral Fellowship to A.A.C., and an NSERC Postdoctoral Fellowship to J.H.A.

Supplementary data

Supplementary data available at *ICB* online.

References

Adams DC, Collyer ML. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Syst Biol 67:14–31.

Adams DC, Otárola-Castillo E. 2013. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods Ecol Evol 4:393–9.

Aguirre LF, Herrel A, van Damme R, Matthysen E. 2002. Ecomorphological analysis of trophic niche partitioning in a tropical savannah bat community. Proc R Soc Lond B Biol Sci 269:1271–8.

Arbour JH, Curtis AA, Santana SE. 2019. Adaptive shifts in skull shape evolution in bats: signatures of echolocation and dietary ecology. Nat Commun 10:2036.

Arnold SJ. 1983. Morphology, performance and fitness. Am Zool 23:347–61.

Bardua C, Wilkinson M, Gower DJ, Sherratt E, Goswami A. 2019. Morphological evolution and modularity of the caecilian skull. BMC Evol Biol 19:1–23.

Bastide P, Ane C, Robin S, Mariadassou M. 2018. Inference of adaptive shifts for multivariate correlated traits. Syst Biol 67:662–80.

Buser TJ, Sidlauskas BL, Summers AP. 2018. 2d or not 2d? testing the utility of 2D vs. 3D landmark data in geometric morphometrics of the sculpin subfamily Oligocottinae (Pisces; Cottoidea). Anat Rec 301:806–18.

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164:683–95.

Cardini A. 2014. Missing the third dimension in geometric morphometrics: how to assess if 2D images really are a good proxy for 3D structures? Hystrix 25:73–81.

Close RI. 1972. Dynamic properties of mammalian skeletal muscles. Physiol Rev 52:129–97.

Cox PG, Fagan MJ, Rayfield EJ, Jeffery N. 2011. Finite element modelling of squirrel, guinea pig and rat skulls: using geometric morphometrics to assess sensitivity. J Anat 219:696–709.

Cox PG, Faulkes CG. 2014. Digital dissection of the masticatory muscles of the naked mole-rat, *Heterocephalus glaber* (Mammalia, Rodentia). PeerJ 2:e448.

Curtis AA, Simmons NB. 2017. Unique turbinal morphology in horseshoe bats (Chiroptera: Rhinolophidae). Anat Rec 300:309–25.

Davis JL, Santana SE, Dumont ER, Grosse I. 2010. Predicting bite force in mammals: two-dimensional versus three-dimensional lever models. J Exp Biol 213:1844–51.

Dickinson E, Stark H, Kupczik K. 2018. Non-destructive determination of muscle architectural variables through the use of diceCT. Anat Rec 301:363–77.

Dumont ER, Davis JL, Grosse IR, Burrows AM. 2011. Finite element analysis of performance in the skulls of marmosets and tamarins. J Anat 218:151–62.

Dumont ER, Herrel A. 2003. The effects of gape angle and bite point on bite force in bats. J Exp Biol 206:2117–23.

Felice RN, Tobias JA, Pigot AL, Goswami A, Felice RN. 2019. Dietary niche and the evolution of cranial morphology in birds. Proc Roy Soc B 286:20182677.

Figueirido B, Tseng ZJ, Martín-Serra A. 2013. Skull shape evolution in durophagous carnivorans. Evolution 67:1975–93.

Ford KL, Albert JS, Evans KM, Kolmann MA, Bernt MJ. 2018. Why the long face? Static allometry in the sexually dimorphic phenotypes of Neotropical electric fishes. Zool J Linn Soc zly076:1–17.

Gehan MA, Hoyer JS, Lorence A, Abbasi A, Hodge JG, Lin A, Doust AN, Feldman MJ, Lizárraga C, Chavez L, et al. 2017. PlantCV v2: image analysis software for high-throughput plant phenotyping. PeerJ 5:e4088.

Gignac PM, Kley NJ. 2014. Iodine-enhanced micro-CT imaging: methodological refinements for the study of the soft-tissue anatomy of post-embryonic vertebrates. J Exp Zool B Mol Dev Evol 322:166–76.

Gignac PM, Kley NJ, Clarke JA. 2016. Diffusible iodine-based contrast-enhanced computed tomography (diceCT): an emerging tool for rapid, high-resolution, 3-D imaging of metazoan soft tissues. J Anat 228:889–909.

Gignac PM, Kley NJ. 2019. 780-sample repeated-measures study to improve visualization of vertebrate soft-tissue anatomy using DiceCT imaging. Society for Integrative and Comparative Biology meeting, Tampa, FL.

Hartstone-Rose A, Perry JMG, Morrow CJ. 2012. Bite force estimation and the fiber architecture of felid masticatory muscles. Anat Rec (Hoboken) 295:1336–51.

Herrel A, De Smet A, Aguirre LF, Aerts P. 2008. Morphological and mechanical determinants of bite force in bats: do muscles matter? J Exp Biol 211:86–91.

Herring SW, Herring SE. 1974. The superficial masseter and gape in mammals. Am Nat 108:561–76.

Herzog W. 1994. Muscle. In: Nigg B, Herzog W, editors. Biomechanics of the musculoskeletal system Chichester: John Wiley & Sons. p. 154–87.

Horn JL. 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30:179–85.

Hylander WL. 1975. The human mandible: lever or link? Am J Phys Anthropol 43:227–42.

Irschick DJ, Higham TE. 2016. Animal athletes: an ecological and evolutionary approach. Oxford: Oxford University Press.

Jeffery NS, Stephenson RS, Gallagher JA, Jarvis JC, Cox PG. 2011. Micro-computed tomography with iodine staining resolves the arrangement of muscle fibres. J Biomech 44:189–92.

Khabbazian M, Kriebel R, Rohe K, Ané C. 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods Ecol Evol 7:811–24.

Linde-Medina M, Boughner JC, Santana SE, Diogo R. 2016. Are more diverse parts of the mammalian skull more labile? Ecol Evol 6:2318–24.

Montgomery K, van de Rijn M, Natkunam Y, Liu CL, Gilks CB, Prapong W, Alizadeh A. 2011. Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. Am J Pathol 161:1557–65.

Olsen AM, Westneat MW. 2015. StereoMorph: an R package for the collection of 3D landmarks and curves using a stereo camera set-up. Methods Ecol Evol 6:351–6.

Openshaw GH, D'Amore DC Vidal-García M, Scott Keogh J. 2017. Combining geometric morphometric analyses of multiple 2D observation views improves interpretation of evolutionary allometry and shape diversification in monitor lizard (*Varanus*) crania. Biol J Linn Soc Lond 120:539–52.

Peres-Neto PR, Jackson DA, Somers KM. 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Comput Stat Data Anal 49:974–97.

Peterson T, Müller GB. 2018. Developmental finite element analysis of cichlid pharyngeal jaws: quantifying the generation of a key innovation. PLoS One 13:1–23.

Phillips JE, Ji L, Rivelli MA, Chapman RW, Corboz MR. 2009. Three-dimensional analysis of rodent paranasal sinus cavities from X-ray computed tomography (CT) scans. Can J Vet Res 73:205–11.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217–23.

Robinson C, Terhune CE. 2017. Error in geometric morphometric data collection: combining data from multiple sources. Am J Phys Anthropol 164:62–75.

Santana SE. 2018. Comparative anatomy of bat jaw musculature via diffusible iodine-based contrast-enhanced computed tomography. Anat Rec 301:267–78.

Santana SE. 2016. Quantifying the effect of gape and morphology on bite force: biomechanical modeling and in vivo measurements in bats. Funct Ecol 30:557–65.

Santana SE, Dumont ER, Davis JL. 2010. Mechanics of bite force production and its relationship to diet in bats. Funct Ecol 24:776–84.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671.

Slater GJ, Dumont ER, Van Valkenburgh B. 2009. Implications of predatory specialization for cranial form and function in canids. J Zool 278:181–8.

Sykes D, Pedroso I, Bernhardt G, Gill RJ, Ahmed F, Abel RL, Raine NE, Smith DB. 2016. Exploring miniature insect brains using micro-CT scanning techniques. Sci Rep 6:1–10.

Uyeda JC, Caetano DS, Pennell MW. 2015. Comparative analysis of principal components can be misleading. Syst Biol 64:677–89.

Wainwright PC. 1994. Functional morphology as a tool in ecological research. In: Wainwright PC, Reilly SM, editors. Ecological morphology. Chicago (IL): The University of Chicago Press. p. 42–59

Watson PJ, Gröning F, Curtis N, Fitton LC, Herrel A, McCormack SW, Fagan MJ. 2014. Masticatory biomechanics in the rabbit: a multi-body dynamics analysis. J R Soc Interface 11:20140564.

Williams SH, Peiffer E, Ford S. 2009. Gape and bite force in the rodents *Onychomys leucogaster* and *Peromyscus maniculatus*: does jaw-muscle anatomy predict performance? J Morphol 270:1338–47.

Xia L, Yang Q, Ge D, Huang C, Lu X. 2013. Geometric morphometric study of the skull shape diversification in Sciuridae (Mammalia, Rodentia). Integr Zool 9:231–45.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004. Geometric morphometrics for biologists: a primer. London: Elsevier Academic Press.