Low resolution scans provide a sufficiently accurate, cost- and time-effective alternative to high resolution scans for interspecific 3D shape analyses (#25642)

First submission

Editor guidance

Please submit by **27 Mar 2018** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Raw data check
Review the raw data. Download from the location described by the author.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files
Download and review all files from the materials page.

7 Figure file(s)
7 Table file(s)
4 Raw data file(s)
2 Other file(s)

Structure and Criteria

2

Structure your review

The review form is divided into 5 sections.
Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Low resolution scans provide a sufficiently accurate, cost- and time-effective alternative to high resolution scans for interspecific 3D shape analyses

Ariel E Marcy [Corresp., 1], **Carmelo Fruciano** [2], **Matthew J Phillips** [3], **Karine Mardon** [4, 5], **Vera Weisbecker** [1]

[1] School of Biological Sciences, University of Queensland, Brisbane, Queensland, Australia

[2] Institut de biologie de l'Ecole normale supérieure, Ecole normale supérieure, Université Paris, Paris, France

[3] School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

[4] Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia

[5] National Imaging Facility, University of Queensland, Brisbane, Queensland, Australia

Corresponding Author: Ariel E Marcy
Email address: a.marcy@uq.edu.au

Background. Advances in three-dimensional (3D) shape capture technology have made powerful shape analyses, such as geometric morphometrics, more feasible. While the highly accurate micro-computed tomography (µCT) scanners have been the "gold standard," recent improvements in 3D surface scanner resolution may make this technology a faster, more portable, and cost-effective alternative. Several studies have already compared the two scanning devices but all use relatively large specimens such as human crania. Here we perform shape analyses on Australia's smallest rodent species to test whether a 3D surface scanner produces similar results to a µCT scanner.

Methods. We captured 19 delicate mouse crania with a µCT scanner and a 3D surface scanner for geometric morphometrics. We ran multiple Procrustes ANOVAs to understand how variation due to scan device compared to other sources of variation such as biologically relevant sources and operator error. We quantified operator error with morphological disparity and repeatability. Finally, we tested whether the different scan datasets could detect intra-specific variation using cross-validation classification. Shape patterns were visualized with Principal Component Analysis (PCA) plots.

Results. In all Procrustes ANOVAs, regardless of factors included, differences between individuals contributed the most to total variation. This is also reflected in the way individuals disperse on the PCA plots. Including only the symmetric component of shape increased the biological signal relative to variation due to device and due to error. 3D scans create a higher level of operator error as evidenced by a greater spread of their replicates on the PCA, a higher morphological disparity, and a lower repeatability score. However, in the test for small intra-specific differences, the 3D scan and µCT scan datasets performed identically.

Discussion. Compared to µCT scans, we find that even very low resolution 3D scans of very small specimens are sufficiently accurate to capture variation at the level of interspecific differences. We also make three recommendations for best use of low resolution data. First, we recommend analyzing the symmetric component of shape to decrease signal from operator error. Second, using 3D scans generates more random error due to increased landmarking difficulty, therefore be conservative in landmark choice and avoid multiple operators. Third, using 3D scans introduces a source of systematic error relative to µCT scans, therefore do not combine them when possible and especially in studies with little variation. Our findings support increased use of low resolution 3D images for most morphological

studies; they are likely applicable to low resolution scans of large specimens made in a medical CT scanner, for example. As most vertebrates are relatively small, we anticipate our results to bolster more researchers designing affordable large scale studies on small specimens with 3D surface scanners.

Low resolution scans provide a sufficiently accurate, cost- and time-effective alternative to high resolution scans for interspecific 3D shape analyses

Authors: Ariel E. Marcy[1], Carmelo Fruciano[2], Matthew J. Phillips[3], Karine Mardon[4,5], Vera Weisbecker[1]

1 School of Biological Sciences, University of Queensland, Brisbane, Australia

2 Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Université Paris, Paris, France

3 School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, Australia

4 Centre for Advanced Imaging, University of Queensland, Brisbane, Australia

5 National Imaging Facility, University of Queensland, Brisbane, Australia

Corresponding Author:

Ariel E. Marcy

a.marcy@uq.edu.au

Abstract

Background. Advances in three-dimensional (3D) shape capture technology have made powerful shape analyses, such as geometric morphometrics, more feasible. While the highly accurate micro-computed tomography (μCT) scanners have been the "gold standard," recent improvements in 3D surface scanner resolution may make this technology a faster, more portable, and cost-effective alternative. Several studies have already compared the two scanning devices but all use relatively large specimens such as human crania. Here we perform shape analyses on Australia's smallest rodent species to test whether a 3D surface scanner produces similar results to a μCT scanner.

Methods. We captured 19 delicate mouse crania with a μCT scanner and a 3D surface scanner for geometric morphometrics. We ran multiple Procrustes ANOVAs to understand how variation due to scan device compared to other sources of variation such as biologically relevant sources and operator error. We quantified operator error with morphological disparity and repeatability. Finally, we tested whether the different scan datasets could detect intra-specific variation using cross-validation classification. Shape patterns were visualized with Principal Component Analysis (PCA) plots.

Results. In all Procrustes ANOVAs, regardless of factors included, differences between individuals contributed the most to total variation. This is also reflected in the way individuals disperse on the PCA plots. Including only the symmetric component of shape increased the biological signal relative to variation due to device and due to error. 3D scans create a higher level of operator error as evidenced by a greater spread of their replicates on the PCA, a higher morphological disparity, and a lower repeatability score. However, in the test for small intra-specific differences, the 3D scan and μCT scan datasets performed identically.

Discussion. Compared to μCT scans, we find that even very low resolution 3D scans of very small specimens are sufficiently accurate to capture variation at the level of interspecific differences. We also make three recommendations for best use of low resolution data. First, we recommend analyzing the symmetric component of shape to decrease signal from operator error. Second, using 3D scans generates more random error due to increased landmarking difficulty, therefore be conservative in landmark choice and avoid multiple operators. Third, using 3D scans introduces a source of systematic error relative to μCT scans, therefore do not combine them when possible and especially in studies with little variation. Our findings support increased use of low resolution 3D images for most morphological studies; they are likely applicable to low resolution scans of large specimens made in a medical CT scanner, for example. As most vertebrates are relatively small, we anticipate our results to bolster more researchers designing affordable large scale studies on small specimens with 3D surface scanners.

Introduction

An organism's shape reveals many facets of its biology, including its evolution, ecology, and functional morphology. In the past three decades, geometric morphometrics has revolutionized the field of shape research with better analysis and visualization of shape complexity (Rohlf & Marcus 1993; Zelditch et al. 2012). As imaging technology continues to advance, three-dimensional (3D) data have become extremely common in geometric morphometric studies, especially in the cases in which 2D data poorly represent the actual 3D object (Buser et al. 2017; Cardini 2014; Fruciano 2016; Reig 1996). 3D capture methods include very high resolution yet high cost and time-intensive options like micro-computed tomography (μCT) scanning. In contrast, 3D surface scanning offers lower acquisition costs and faster scanning, but has the

disadvantage of generally lower resolution, which limits its use on very small specimens (Fig. 1). For confident use of surface scans in small specimens, it is therefore important to assess the measurement error introduced by choosing a 3D surface scanner for geometric morphometrics.

Most vertebrates would be considered small, for example about two thirds of mammals are below 10kg (Weisbecker & Goswami 2010), which would translate to small skeletal specimens. Therefore, morphometric studies proposing large sample sizes must be very well funded to use a μCT scanner or have a low-cost option, such as a 3D surface scanner. Previous studies have compared μCT scans to 3D surface scans, however, these were all done in large animals, primarily primates (Badawi-Fayad & Cabanis 2007; Fourie et al. 2011; Katz & Friess 2014; Robinson & Terhune 2017; Sholts et al. 2010; Slizewski et al. 2010). While these studies found low error and high repeatability in 3D surface scans similar to μCT scans, there was a suggestion that higher error occurred in the sample's smaller specimens (Badawi-Fayad & Cabanis 2007; Fourie et al. 2011). Other recent studies have conducted 3D geometric morphometric studies on small vertebrate skulls but nearly all have relied exclusively on μCT scanning (Cornette et al. 2013; Evin et al. 2011). The only exception we are aware of is Munoz-Munoz et al. (2016), which successfully used photogrammetry – a technique combining 2D photographs into a 3D model – to analyze domestic mouse skulls (*Mus musculus domesticus,* C Linnaeus, 1758). Photogrammetry, like 3D surface scanning, is a low-cost alternative to μCT and comes with its own trade-offs in time and scan resolution (Katz & Friess 2014). Compared to the new generation of blue light surface scanners, photogrammetry requires more time for image acquisition and for file processing (Katz & Friess 2014). A previous study on a single macaque specimen reported inconsistent levels of error across operators and scanners, which contributed

to the lack of general pattern for differences across scanners/resolutions (Shearer et al. 2017). However, using an interspecific dataset, Fruciano et al. (2017) reported higher repeatability for the higher resolution scans and 2.07-11.26% of total variance due to scan type (depending on device, operator and landmark set combination). We expect that small specimens would exacerbate any variation due to device and the interaction of device with other factors, such as landmark choice and operator. More work comparing these different methods – μCT scanning, 3D surface scanning, and photogrammetry – will allow researchers to make an informed decision. For example, for those with time constraints in museum collections, a fast 3D surface scanner may be the best option if the resolution is suitable for specimen size.

The lower resolution of 3D surface scanners may increase both random and systematic measurement error, which is exacerbated by small specimens because operators may have more difficulty identifying landmark locations (Arnqvist & Martensson 1998; Fruciano 2016). Random error increases variance without changing the mean; this "noise" dilutes biologically informative patterns and, in principle, decreases statistical power (Arnqvist & Martensson 1998; Fruciano 2016). By contrast, systematic error is non-randomly distributed, thus changing the mean and introducing bias to the data (Arnqvist & Martensson 1998; Fruciano 2016). Error assessment can be done with repeated measures of the same individuals (e.g. Fruciano et al. 2017; Munoz-Munoz & Perpinan 2010; Robinson & Terhune 2017) or by comparison to a "gold standard" or ideal representation of the specimens (Fruciano 2016; Slizewski et al. 2010; Williams & Richtsmeier 2003) such as can be achieved with a high resolution μCT scan. Repeated measure designs can uncover this systematic error, for example, if one 3D capture method differs from another in a specific, non-random, pattern (Fruciano 2016; Fruciano et al.

2017). Furthermore, designs including repeated measures of the same individuals allow partitioning of variance into components, quantifying error due to scan type as compared to biologically-relevant sources of variation such as asymmetry (Fruciano 2016; Klingenberg et al. 2002; Klingenberg & McIntyre 1998).

In this study, we quantify the error introduced by studying specimens of a size at the very lower limits of surface scanner resolution. This situation could also arise when using relatively large specimens, which are nonetheless at the lower limit of a medical CT scanner's resolution for example. We test whether the complex shape of very small specimens can be adequately captured using an HDI109 3D surface scanner with a stated resolution of 80 µm as compared to a µCT scanner with a resolution of 28 µm. To do so, we use the delicate mouse (*Pseudomys delicatulus*, J Gould, 1842), one of the smallest rodents in the world with a 55-75 mm head-and-body length (Breed & Ford 2007). The miniscule *P. delicatulus* crania (~20mm) are at the edge of the HDI109 3D surface scanner's range thus providing an extreme test of this scanning method (Fig. 1, Fig. 2).

Methods

Data collection

We selected 19 adult individuals, male and female, of *Pseudomys delicatulus* from the Queensland Museum in Brisbane, Australia (specimen numbers and sexes in Additional File 1: Table S1). The cranium from each individual was scanned at the Centre for Advanced Imaging at the University of Queensland in a µCT scanner (Siemens Inveon PET/CT scanner). The scanner was operated at 80 KV energy, 250 µA intensity with 540 projections per 360°, a medium-high

133 magnification with bin 2 was applied, and 2000 ms exposure time. The samples were scanned at
134 a nominal isotropic resolution of 28 µm. The data were reconstructed using a Feldkamp
135 conebeam back-projection algorithm provided by an Inveon Acquisition workstation from
136 Siemens (IAW version 2.1). Surface models were obtained using Mimics Research version 20.0.
137
138 Each cranium was also scanned by 3D LMI's HDI109 blue light surface scanner with a
139 resolution of 80 µm. For brevity, we will refer to this method as 3D scanning. For this method,
140 the cranium was placed on a rotary table providing the scanner with 360 views. To capture the
141 entire shape, the cranium was scanned in three different orientations: one ventral view with the
142 cranium resting on the frontals and two dorsal views with the cranium tipped to each side, resting
143 on an incisor, auditory bulla, and zygomatic arch. To assist others in replicating our HDI109 3D
144 surface scanning on small specimens, we have included a Standard Operating Procedure with our
145 settings (Additional File 2: Supplementary Methods).
146
147 We duplicated the digital file for each unique individual-scan method combination three times
148 such that each individual was represented by 6 replicates, giving a total sample of 114 replicates
149 (Fig. 2a). Each replicate was landmarked in Viewbox version 4.0 (dHAL software, Kifissia,
150 Greece; www.dhal.com; Polychronis et al. 2013). To capture shape, we placed 58 fixed
151 landmarks, 145 sliding semi-landmarks, and 86 sliding patch points (3D meshes defined by
152 semi-landmark borders) for a total of 289 points (Fig. 3, Additional File 3: Table S2). We used
153 the template feature in Viewbox to semi-automate the placement of semi-landmark curves and to
154 fully automate the placement of patch points. Our landmark design covered most important
155 biological structures except for the zygomatic arch (Fig. 3); we avoided this fine structure

because dehydration and loss of support from surrounding muscles during skeletonization almost certainly causes specimen preparation error (Schmidt et al. 2010; Yezerinac et al. 1992).

Data analysis

The landmark coordinates for all 114 replicates were aligned using a generalized Procrustes superimposition implemented in the R package *geomorph* (v. 3.0.5) (Adams 2016; Adams & Otarola-Castillo 2013). Superimposition of each set of landmark coordinates removes differences in size, position, and orientation, leaving only shape variation (Rohlf & Slice 1990). Semi-landmarks and patches were permitted to slide along their tangent directions to minimize Procrustes distance between replicates (Gunz et al. 2005). The resulting Procrustes tangent coordinates were used as shape variables in all subsequent shape analyses. All our statistical analyses were performed either in R (v. 3.3.3) using the R packages *geomorph* (v. 3.0.5) (Adams 2016; Adams & Otarola-Castillo 2013) and *Morpho* (v. 2.5.1) (Schlager 2017) or using MorphoJ (v. 1.06d) (Klingenberg 2011).

First, asymmetry is a known source of variation within a sample (Klingenberg et al. 2002), so we tested for it with MorphoJ's general Procrustes ANOVA function and subsequently removed it (Fig. 2b). Isolating symmetric shape has been done in other 3D surface scanner studies where operator and device error have been of the same magnitude as asymmetric error (Fruciano et al. 2017). Variation due to asymmetry is more impacted by operator error because of its smaller effect sizes compared to variation among individuals (Fruciano 2016; Fruciano et al. 2017; Klingenberg et al. 2010; Leamy & Klingenberg 2005). This suggests that low resolution studies on asymmetry would be negatively impacted. For this reason, we performed all subsequent

179 analyses on the symmetric shape component. We then performed a PCA on the symmetric shape
180 variables to visualize the variation between individuals, within scan method replicates, and
181 between scan method replicates. As an exploratory analysis, PCA can help intuitively visualize
182 both random error (greater spread of one scan method replicate compared to the other) and
183 systematic error (repeated pattern of one scan method shifting relative to another). However,
184 further analyses are necessary to quantify these sources of error.
185
186 Second, our replicate design allowed us to assess whether an operator digitizing one type of scan
187 was more variable in landmark placement than when digitizing scans from the other device (Fig.
188 2c). We did so by computing the Procrustes variance for each individual/device combination. In
189 *geomorph*, Procrustes variances are calculated for each set of observations (i.e. replicates) as the
190 sum of the diagonal elements of the set's covariance matrix divided by the number of
191 observations (Adams 2016; Zelditch et al. 2012). We computed Procrustes variance for each
192 combination of individual and device so that Procrustes variance reflects only variation due to
193 digitization. We then compared Procrustes variance between devices using a box plot and the
194 permutational procedure implemented in *geomorph*. Next we quantified digitization consistency
195 by computing repeatability (i.e. the intraclass correlation coefficient using the Procrustes
196 ANOVA mean squares) for each device as suggested by Fruciano (2016). This value is normally
197 comprised between 0 and 1, with values close to 1 indicating much larger variation due to the
198 factor used in computing the Procrustes ANOVA (in our case, variation among individuals)
199 compared to residual variation (in our case, variation among digitizations). In other words,
200 comparing repeatability between devices gives a similar information to the one obtained by the
201 box plots of Procrustes variance but on a more easily interpretable scale from 0 to 1.

202

203 Finally, we investigated whether there is a difference between devices in a commonly used shape
204 analysis: the detection and correct classification of sexual dimorphism (Fig. 2c). We began with
205 a Procrustes ANOVA in R on the symmetric component for the subset of individuals with sex
206 information (n = 11 distinct individuals; n = 66 replicates). This allowed us to gauge the
207 magnitude of the effect of sexual dimorphism compared to other sources of variation, as well as
208 test for significant differences in mean shape between males and females. Then with *Morpho*, we
209 averaged the shape of each replicate triad for each device, performed a between group PCA
210 using sex as group and then a cross-validation of classification accuracy (Schlager 2017).

211

212 **Results**

213 **Analyses of shape variation**

214 Our Procrustes ANOVA results indicate that variation among individuals (%Var = 47.4)
215 contributes the most, with asymmetry (fluctuating and directional), device, and operator error
216 contributing the remainder, in order of greatest to least (Table 1). The %Var values indicate that
217 directional asymmetry contributes a similar amount of variation as other sources of non-
218 biological variation and that fluctuating asymmetry accounts for much less than digitization error
219 and variation between devices (Table 1). This means that using analyses of asymmetry using a
220 combination of μCT and 3D surface scans would likely be unreliable in specimens the size of
221 delicate mice or for specimens scanned at a similarly low resolution. Furthermore, since
222 digitization error is large compared to the components of asymmetric variation, even a single
223 device yet low resolution study of asymmetry would likely be unreliable unless appropriate
224 arrangements are made to reduce error (Fruciano 2016).

225

226 The Procrustes ANOVA on the symmetric component of shape reports the individual shape
227 representing biological variation is 73.3% (Table 2). Differences between scan devices represent
228 14.5% and the residuals encompassing differences among replicates or operator error represent
229 12.2% of total variance (Table 2). Thus, our Procrustes ANOVA shows that most of the variation
230 is represented by biological variation but the significance of the variation due to device may
231 indicate systematic error.

232

233 The PCA of our symmetric dataset revealed that the first 3 principal components (PCs) account
234 for 47.1% of total variation (PC1 = 26.4%, PC2 = 11.9%, PC3 = 8.9%, n = 114) (Fig. 4). Each of
235 the remaining PCs accounted for 5% or less of total variation therefore we only considered the
236 first three. Positive values along PC1 correspond to a larger braincase relative to the rostrum
237 (Fig. 5a). Positive values along PC2 correspond to a wider frontal bone (Fig. 5b). Finally,
238 positive values along PC3 correspond to a more convex, dorsally-curved ventral surface (Fig.
239 5c).

240

241 The plot of PC1 and PC2 supports the results from the symmetric Procrustes ANOVA in that
242 most of the visible variation is between clusters of each individual's replicates. Indeed,
243 regardless of scanning device, replicates from the same individual cluster together (Fig. 4a). For
244 most individuals, replicates occupy non-overlapping morphospaces except for those around the
245 crowded mean shape (Fig. 4a). Within each individual's morphospace, μCT replicates usually
246 form a tighter cluster than the 3D replicates (Fig. 4a). This pattern suggests that using μCT scans
247 introduces less random error than using 3D scans. Furthermore, within an individual, 3D scan

replicates tend to cluster closer to other 3D replicates while μCT scan replicates tend to cluster closer to other μCT replicates (Fig. 4a). This supports the interpretation for a systematic difference between scan method shape means reported the Procrustes ANOVA's significant scan variation component (Table TK). Indeed, for most individuals, 3D scan replicates score higher than their μCT scan replicates on both PC1 and PC2. This suggests the systematic error may be driven by 3D scans overestimating both braincase volume and frontal bone width relative to μCT scans (Fig. 4a, Fig. 5a,b).

Overall, plots of the scores along the first two components mirror and provide intuitive visualization to the patterns observed in the analyses using Procrustes ANOVA. The plot of PC1 and PC3 highlights another possible systematic difference between 3D and μCT scans (Fig. 4b). The PC3 axis displaces μCT replicates from 3D replicates such that individuals no longer occupy distinct morphospaces (Fig. 4b). On the PC3 axis, μCT scan replicates consistently score higher, which corresponds to a more dorsally curved ventral surface relative to 3D scan replicates (Fig. 4b, Fig. 5c). Along with PC1 and PC2, PC3's result strengthens the signal for a general pattern of a difference in the degree of surface curvature captured by 3D and μCT scanners, which could be contributing to the systematic error reported by the Procrustes ANOVA (Table 2). In summary, despite a small but morphologically significant source systematic error, both the Procrustes ANOVA and the PCA report that most variation comes from a biological signal, the differences between individuals.

Analyses of variance and error

To compare the digitization error in each scanning device dataset, we calculated the Procrustes variance among the replicate triads of each individual. We found that Procrustes variance is significantly ($p<0.001$) higher in 3D scans ($1.34x10^{-4}$) than μCT ($4.81x10^{-5}$) scans (Fig. 6). This means that digitizations are more variable in 3D scans than in μCT which is consistent with decreased clustering in 3D scans relative to μCT scans in the PCAs (Fig. 4).

The repeatability scores for each scan dataset mirrored the Procrustes variance results but with a more intuitive number on a 0-1 scale. We found that the μCT scan dataset had a repeatability of 0.927 and the 3D scan data had a repeatability of 0.814 (Table 3). This means operators have an easier time repeating their digitizations (i.e. landmark placements) with μCT scans than with 3D scans.

Analyses with a biological example: sexual dimorphism

A subset of our dataset had sex information (n = 11; f = 7, m = 4), allowing us to perform a test on whether using different scan devices to detect a very subtle intra-specific signal produces different results. Our symmetric Procrustes ANOVA on individuals, sex, and device found that differences between individuals is still the largest component (Table 4; Rsq = 0.691) with variation due to device (Rsq = 0.172) and sex/residuals (Rsq = 0.137) contributing similar amounts. Variation due to device is larger than variation due to sex, which suggests that 3D scans and μCT scans should not be combined for similar analyses. However, the between group PCAs do not suggest marked sexual dimorphism to begin with plots (Fig. 7). Therefore, the subtly of this biological signal could be the main reason for the relatively low contribution of sex to total variation. Finally, we performed a cross-validation test on the between group PCAs to

293 assess which scan dataset can more reliably identify sexes based on shape (Table 5). The results
294 show that in this case, 3D scans and μCT scans perform identically (overall classification
295 accuracy = 64%).
296
297 **Discussion**
298 In this study, we contrasted very high resolution μCT scans with their extreme opposite: 3D
299 surface scans of very small specimens. Our low versus high resolution datasets allowed us to
300 assess whether the low resolution scans still allow defensible investigations of biological shape
301 variation. We found that despite the low quality of the 3D scans, sufficient amounts of biological
302 variation are present to perform, at the very least, interspecific comparisons. In datasets with
303 only very slight intra-specific differences does the ability to distinguish biological signal from
304 error's "noise" occur. For example, the subtle sexual dimorphism in our small sample was only
305 just detected. However, we present three considerations to make before using low resolution
306 datasets. First, we found that we needed to remove the signal from asymmetry to investigate
307 shape variation more confidently. This makes low resolution datasets a poor choice for studies
308 on asymmetry. Second, using 3D scans creates more random error due to increased landmarking
309 difficulty, therefore care should be taken in landmark choice, and possibly landmarking software
310 and operator choice. Digitization error may also be reduced by taking averages of repeated
311 measurements (Arnqvist & Martensson 1998; Fruciano 2016). Third, using 3D scans also
312 introduces a source of systematic error relative to μCT scans, therefore we recommend not
313 combining them whenever possible (see also Fruciano et al. 2017), and especially in studies on
314 small intra-specific variation. In summary, with a few precautions listed above, we expect that

for studies with similarly sized skulls or similarly low resolution scans, the variation due to error will be sufficiently low for successful detection of interspecific shape differences.

Measurement error and 3D scan reliability

Systematic error between the two scan devices is shown by consistent displacement patterns in the PCA. Indeed, across all three PC axes, the scans differ in how they measure concavity around the braincase, frontal, and ventral surface. This systematic pattern could suggest that the 3D scanner technology errs on adding volume to the digital specimen relative to the μCT scan but it could also be the other way around with the μCT scan distorting the images. Furthermore, even when using the symmetric component of shape, the percent of variation contributed by scan device is quite substantial at about 14.5%. Because scan device contributes this much to variation and because systematic error between scan device exists, researchers expecting very small variation due to biological sources would be advised not to combine 3D scan and μCT scan datasets. However, overall each individual's 3D and μCT replicates almost always occupied distinct areas of the morphospace, supporting their comparability for most morphometric studies.

While the two scan methods are usually comparable, using the low resolution 3D scans introduces more digitization error than the higher resolution μCT scans, which likely reflects increased user error due to lower resolution in 3D scans. This increased random error is reflected in both the larger point clouds of 3D replicates relative to μCT replicates in the PCAs as well as the higher morphological disparity and lower repeatability score of 3D scans. As expected, we found that the low resolution 3D scans were more difficult to landmark because key cranial features such as sutures and smaller processes were less distinct (Fig. 1 versus Fig. 3).

Nevertheless, our 3D scan repeatability score of 0.82 appears consistent with the literature: it is much lower than 3D scanned human-sized skulls – above 0.95 (Badawi-Fayad & Cabanis 2007; Fourie et al. 2011) but it is within the range of 3D scanned macropodoids (e.g., kangaroos) – 0.78-0.98, depending on device and landmark choice (Fruciano et al. 2017). This trend of decreasing repeatability with decreasing body size may reflect measurement error becoming a larger percentage of overall size (Robinson & Terhune 2017). Relatedly, recent work has shown that unreliable landmarks, or those with greater variability in placement, significantly decrease repeatability (Fruciano et al. 2017). This may be especially true for small specimens, for which small variations from the landmark location represent a larger percentage of their overall size.

This study did not look at multiple operator error which can be considerable, particularly if difficult landmarks are included (Fruciano et al. 2017). If inter-operator error were combined with the resolution-driven measurement error found here, it is possible that biological signal would diminish to a degree that could not support even interspecific comparisons.

Measurement error compared to biological variation

The challenge of any quantitative measurement study is to minimize measurement error introduced from various sources (in our case, device, resolution, and observer) relative to the "true" signal of biological variation. In the case of inter-observer error, which is one measurement error source, several studies suggest that interspecific variation overwhelms inter-observer such that it does not pose an issue with the correct interpretation of results (Robinson & Terhune 2017).

In our test on the detectability of sexual dimorphism relative to scan device, we showed that while variation contributed by each was similar (and that from scan device slightly higher), both scan datasets detected a small sexually dimorphic pattern and they performed equally. This suggests that 3D scans may even be acceptable for detecting some intra-specific patterns. This was a small sample (n = 11) therefore further study with larger datasets would improve confidence for using 3D scans for intra-specific studies. Nevertheless, it is promising that 3D scans and μCT scans performed similarly even at such a small sample size.

Choosing a digitization method: 3D surface scanning versus μCT versus photogrammetry

With many options for digitizing 3D specimens available, decisions on the acquisition mode must consider price, scanning time, processing time, portability, and scan resolution. The one-off investment of a relatively high resolution 3D surface scanner such as the HDI109 provided a model portable enough to take on airplanes and has fast scanning and processing times. Our model took 10 minutes from starting the scan to the finished surface file, but note that larger specimens requiring multiple sub-scans will take longer. These fast acquisition times are an asset in collection efforts that rely on expensive and time-limited museum travel. For example, one of us (AEM) digitized over 100 individuals in one week using the same scanning protocol. However, the quality and speed of scanning varies by model; for example, other 3D surface scanners could take over 45 minutes to capture one specimen and may also require more effort to process scans (Katz & Friess 2014).

Compared to 3D surface scanners, μCT scanners provide much higher resolution, which in this study translated into less measurement error. However, uCT facilities are not widely accessible,

not mobile, and tend to be more expensive. Depending on the facility, µCT scanning involves transport to the facility, scanning either by the operator, processing scans into image stacks, and finally loading scans into specialized (and frequently high-cost) software to do the 3D reconstruction. These reconstructions can be time consuming especially if the cranium needs to be separated from the mandibles. Finally, specimens need to be loaned from their collections for uCT acquisition, which requires specimen transport and curator permission and is particularly difficult when large numbers of specimens from distant locations need to be scanned.

This study did not investigate photogrammetry, which is another and increasingly popular method for digitizing 3D shape. This method uses software to align 2D photographs taken from many different views into a 3D file. Photogrammetry is much cheaper and more portable than 3D surface scanning since it only requires a camera of suitable resolution and very affordable photo-alignment software like Agisoft PhotoScan (Agisoft LLC, St. Petersburg, Russia; www.agisoft.com). The trade-offs are that in our experience, photogrammetry takes at least three times longer to acquire the photos, it involves higher risk of human error or inconsistency during photography, and it requires an order of magnitude more time to align the photos into a 3D digital file. While photo-alignment can be done at convenience after photography, the greater time required to capture enough photos may be a deciding factor for researchers with time limitations in museum collections. As for scan resolution, photogrammetry may perform better than 3D surface scanners in some cases (Fourie et al. 2011) or at least provide an acceptable alternative (Katz & Friess 2014; Muñoz-Muñoz et al. 2016).

Conclusions

In summary, the best 3D capture method will vary based on the study's design, expected effect size for the biological variation of interest, and the researcher's limitations on time, money, and travel. In addition to image resolution requirements, it is wise to assess the time it will take to capture and process each specimen as well as portability needs. Here, we have shown that a 3D surface scanner can provide an acceptable alternative to a μCT scanner for assessing biological signal of 3D shape even in small specimens that are at the limits of 3D scanner resolution. Furthermore, as previously suggested (e.g., Fruciano 2016), exploratory pilot studies of measurement error are advisable when practically possible. We recommend a preliminary test on multiple devices – including surface scanners – of how levels of error compare to biological signal and whether there is substantial systematic error. Doing so may provide a defensible alternative to an expensive and time consuming large-scale acquisition of μCT scans.

Acknowledgements

We would like to thank Cruise Speck for assistance with Viewbox software and Dr. Heather Janetzki for hosting us in the mammal collections at the Queensland Museum.

Abbreviations

Landmark (LM)

Micro-computed tomography (μCT)

Principal component analysis (PCA)

Principal component (PC)

Three-dimensional (3D)

430

431 **References**

432 Adams D, ML Collyer, and E. Sherratt. 2016. geomorph: Software for geometric morphometric
433 analyses. 3.0 ed.

434 Adams DC, and Otarola-Castillo E. 2013. geomorph: an r package for the collection and analysis
435 of geometric morphometric shape data. *Methods in Ecology and Evolution* 4:393-399.
436 10.1111/2041-210x.12035

437 Arnqvist G, and Martensson T. 1998. Measurement error in geometric morphometrics: Empirical
438 strategies to assess and reduce its impact on measures of shape. *Acta Zoologica*
439 *Academiae Scientiarum Hungaricae* 44:73-96.

440 Badawi-Fayad J, and Cabanis EA. 2007. Three-dimensional procrustes analysis of modern
441 human craniofacial form. *Anatomical Record-Advances in Integrative Anatomy and*
442 *Evolutionary Biology* 290:268-276. 10.1002/ar.20442

443 Breed B, and Ford F. 2007. *Native mice and rats*: CSIRO PUBLISHING.

444 Buser TJ, Sidlauskas BL, and Summers AP. 2017. 2D or Not 2D? Testing the Utility of 2D Vs.
445 3D Landmark Data in Geometric Morphometrics of the Sculpin Subfamily Oligocottinae
446 (Pisces; Cottoidea). *The Anatomical Record*.

447 Cardini A. 2014. Missing the third dimension in geometric morphometrics: how to assess if 2D
448 images really are a good proxy for 3D structures? *Hystrix-Italian Journal of Mammalogy*
449 25:73-81. 10.4404/hystrix-25.2-10993

450 Cornette R, Baylac M, Souter T, and Herrel A. 2013. Does shape co-variation between the skull
451 and the mandible have functional consequences? A 3D approach for a 3D problem.
452 *Journal of Anatomy* 223:329-336. 10.1111/joa.12086

453 Evin A, Horacek I, and Hulva P. 2011. Phenotypic diversification and island evolution of
454 pipistrelle bats (Pipistrellus pipistrellus group) in the Mediterranean region inferred from
455 geometric morphometrics and molecular phylogenetics. *Journal of Biogeography*
456 38:2091-2105. 10.1111/j.1365-2699.2011.02556.x
457 Fourie Z, Damstra J, Gerrits PO, and Ren YJ. 2011. Evaluation of anthropometric accuracy and
458 reliability using different three-dimensional scanning systems. *Forensic Science*
459 *International* 207:127-134. 10.1016/j.forsciint.2010.09.018
460 Fruciano C. 2016. Measurement error in geometric morphometrics. *Development Genes and*
461 *Evolution* 226:139-158. 10.1007/s00427-016-0537-4
462 Fruciano C, Celik MA, Butler K, Dooley T, Weisbecker V, and Phillips MJ. 2017. Sharing is
463 caring? Measurement error and the issues arising from combining 3D morphometric
464 datasets. *Ecology and Evolution* 7:7034-7046. 10.1002/ece3.3256
465 Gunz P, Mitteroecker P, and Bookstein FL. 2005. Semilandmarks in Three Dimensions. *Modern*
466 *Morphometrics in Physical Anthropology*:73-98. 10.1007/0-387-27614-9_3
467 Katz D, and Friess M. 2014. 3D from standard digital photography of human crania—a
468 preliminary assessment. *American Journal of Physical Anthropology* 154:152-158.
469 Klingenberg C, Wetherill L, Rogers J, Moore E, Ward R, Autti-Rämö I, Fagerlund Å, Jacobson
470 S, Robinson L, and Hoyme H. 2010. Prenatal alcohol exposure alters the patterns of
471 facial asymmetry. *Alcohol* 44:649-657.
472 Klingenberg CP. 2011. MorphoJ: an integrated software package for geometric morphometrics.
473 *Molecular Ecology Resources* 11:353-357. 10.1111/j.1755-0998.2010.02924.x
474 Klingenberg CP, Barluenga M, and Meyer A. 2002. Shape analysis of symmetric structures:
475 Quantifying variation among individuals and asymmetry. *Evolution* 56:1909-1920.

476 Klingenberg CP, and McIntyre GS. 1998. Geometric morphometrics of developmental
477 instability: Analyzing patterns of fluctuating asymmetry with procrustes methods.
478 *Evolution* 52:1363-1375. 10.2307/2411306
479 Leamy LJ, and Klingenberg CP. 2005. The genetics and evolution of fluctuating asymmetry.
480 *Annu Rev Ecol Evol Syst* 36:1-21.
481 Munoz-Munoz F, and Perpinan D. 2010. Measurement error in morphometric studies:
482 comparison between manual and computerized methods. *Annales Zoologici Fennici*
483 47:46-56.
484 Munoz-Munoz F, Quinto-Sanchez M, and Gonzalez-Jose R. 2016. Photogrammetry: a useful
485 tool for three-dimensional morphometric analysis of small mammals. *Journal of*
486 *Zoological Systematics and Evolutionary Research* 54:318-325. 10.1111/jzs.12137
487 Muñoz-Muñoz F, Quinto-Sánchez M, and González-José R. 2016. Photogrammetry: a useful tool
488 for three-dimensional morphometric analysis of small mammals. *Journal of Zoological*
489 *Systematics and Evolutionary Research*.
490 Polychronis G, Christou P, Mavragani M, and Halazonetis DJ. 2013. Geometric Morphometric
491 3D Shape Analysis and Covariation of Human Mandibular and Maxillary First Molars.
492 *American Journal of Physical Anthropology* 152:186-196. 10.1002/ajpa.22340
493 Reig S. 1996. Correspondence between interlandmark distances and caliper measurements.
494 *Advances in Morphometrics* 284:371-385.
495 Robinson C, and Terhune CE. 2017. Error in geometric morphometric data collection:
496 Combining data from multiple sources. *American Journal of Physical Anthropology*
497 164:62-75.

498 Rohlf FJ, and Marcus LF. 1993. A REVOLUTION IN MORPHOMETRICS. *Trends in Ecology*
499 *& Evolution* 8:129-132.
500 Rohlf FJ, and Slice D. 1990. Extensions of the Procrustes method for the optimal
501 superimposition of landmarks. *Systematic Zoology* 39:40-59. 10.2307/2992207
502 Schlager S. 2017. Morpho and Rvcg -- Shape Analysis in R. In: Guoyan Zheng SLaGS, ed.
503 *Statistical Shape and Deformation Analysis*: Academic Press, 217--256.
504 Schmidt EJ, Parsons TE, Jamniczky HA, Gitelman J, Trpkov C, Boughner JC, Logan CC,
505 Sensen CW, and Hallgrimsson B. 2010. Micro-computed tomography-based phenotypic
506 approaches in embryology: procedural artifacts on assessments of embryonic craniofacial
507 growth and development. *Bmc Developmental Biology* 10. 10.1186/1471-213x-10-18
508 Shearer BM, Cooke SB, Halenar LB, Reber SL, Plummer J, Delson E, and Tallman M. 2017.
509 Evaluating causes of error in landmark-based data collection using scanners. *Plos One*
510 12:e0187452.
511 Sholts SB, Wärmländer SKTS, Flores LM, Miller KWP, and Walker PL. 2010. Variation in the
512 Measurement of Cranial Volume and Surface Area Using 3D Laser Scanning
513 Technology. *Journal of Forensic Sciences* 55:871-876. 10.1111/j.1556-
514 4029.2010.01380.x
515 Slizewski A, Friess M, and Semal P. 2010. Surface scanning of anthropological specimens:
516 nominal-actual comparison with low cost laser scanner and high end fringe light
517 projection surface scanning systems. *Quartär* 57:179-187.
518 Weisbecker V, and Goswami A. 2010. Brain size, life history, and metabolism at the
519 marsupial/placental dichotomy. *Proceedings of the National Academy of Sciences*
520 107:16216-16221. 10.1073/pnas.0906486107

521 Williams FL, and Richtsmeier JT. 2003. Comparison of mandibular landmarks from computed
522 tomography and 3D digitizer data. *Clinical Anatomy* 16:494-500. 10.1002/ca.10095
523 Yezerinac SM, Lougheed SC, and Handford P. 1992. MEASUREMENT ERROR AND
524 MORPHOMETRIC STUDIES - STATISTICAL POWER AND OBSERVER
525 EXPERIENCE. *Systematic Biology* 41:471-482. 10.2307/2992588
526 Zelditch ML, Swiderski DL, and Sheets HD. 2012. Geometric Morphometrics for Biologists: A
527 Primer, 2nd Edition. *Geometric Morphometrics for Biologists: a Primer, 2nd Edition*:1-
528 478.

Figure 1

Low resolution 3D surface scans of delicate mouse crania.

(A) Dorsal view. (B) Lateral view. (C) Ventral view. See Figure 3 to compare with the much higher resolution of µCT scans. All crania are rendered in Viewbox v. 4.0.

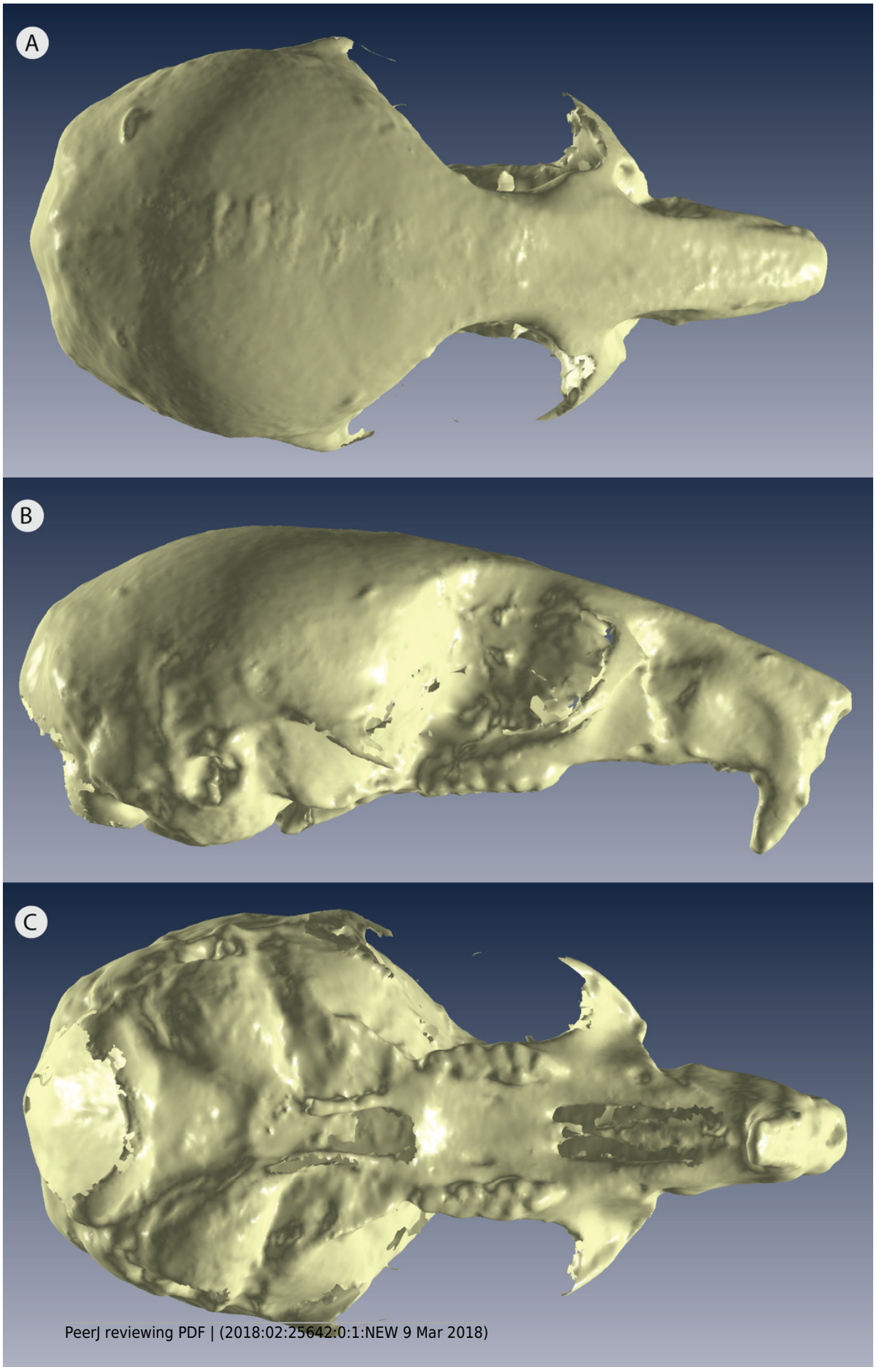

A
B
C

Figure 2

Methods flow diagram highlighting the relationship between our questions and our analyses.

(A) All delicate mouse (*Pseudomys delicatulus*) crania were sourced from the Queensland Museum in Brisbane, Australia. Landmarks (LMs) capture homologous points, semi-landmarks (semi-LMs) capture curves between landmarks, and patch points capture surfaces between landmarks and semi-landmarks. (B - D) These sections of questions and associated figure and table numbers summarize how we organize the paper, particularly the Results, into three sets of related analyses.

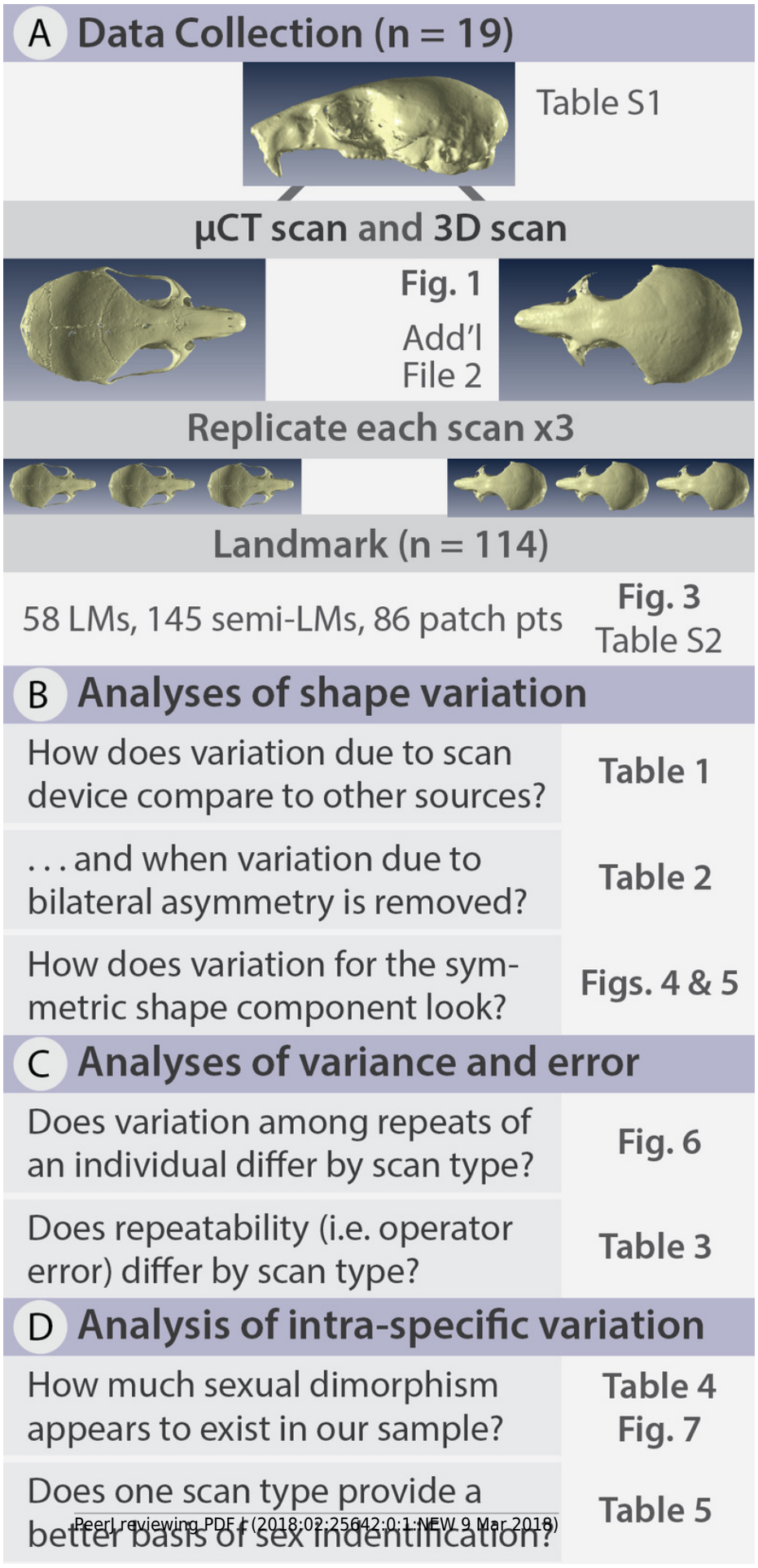

A Data Collection (n = 19)
Table S1
µCT scan and 3D scan
Fig. 1
Add'l File 2
Replicate each scan x3
Landmark (n = 114)
58 LMs, 145 semi-LMs, 86 patch pts
Fig. 3
Table S2
B Analyses of shape variation
How does variation due to scan device compare to other sources?
Table 1
. . . and when variation due to bilateral asymmetry is removed?
Table 2
How does variation for the symmetric shape component look?
Figs. 4 & 5
C Analyses of variance and error
Does variation among repeats of an individual differ by scan type?
Fig. 6
Does repeatability (i.e. operator error) differ by scan type?
Table 3
D Analysis of intra-specific variation
How much sexual dimorphism appears to exist in our sample?
Table 4
Fig. 7
Does one scan type provide a better basis of sex identification?
Table 5

Figure 3

Positions of landmarks for geometric morphometric analyses.

Locations of fixed landmarks (black points), sliding semi-landmarks (red points) and sliding surface patches (purple points) on a µCT scanned individual. (A) Dorsal view of the cranium. (B) Lateral view. (C) Ventral view. Definitions are given in Table S2.

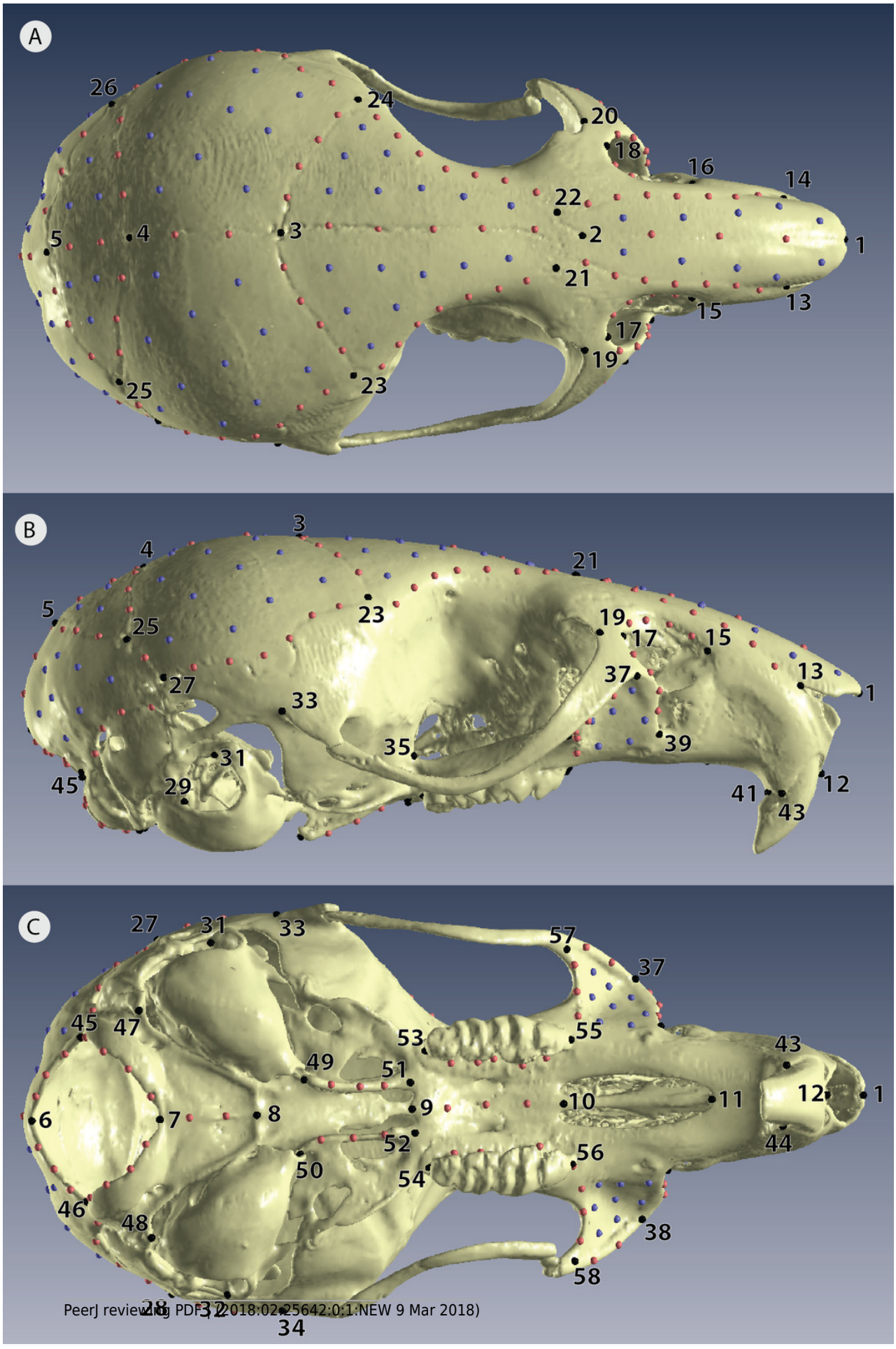

A
26
24
20
18
16
14
22
5
4
3
2
1
21
15
13
17
19
25
23
B
3
4
21
5
23
19
25
17
15
13
27
37
1
33
35
39
31
45
29
41
12
43
C
27
31
33
57
37
45
47
53
55
43
49
51
12
1
6
7
8
9
10
11
52
44
56
50
54
46
48
38
58
28
32
34

Figure 4

Exploratory PCA plots of shape variation showing differences among individuals, scan devices, and replicates of the same scan device.

A) PC1 versus PC2 and **B)** PC1 versus PC3. Each individual has a unique color shared by all of its 6 replicates. Each replicate's point is labeled for its scan device, either "CT" for µCT scanned or "3D" for 3D surface scanned. Each axis reports the total variance explained by that principal component: 26.4% for PC1, 11.9% for PC2, and 8.9% for PC3.

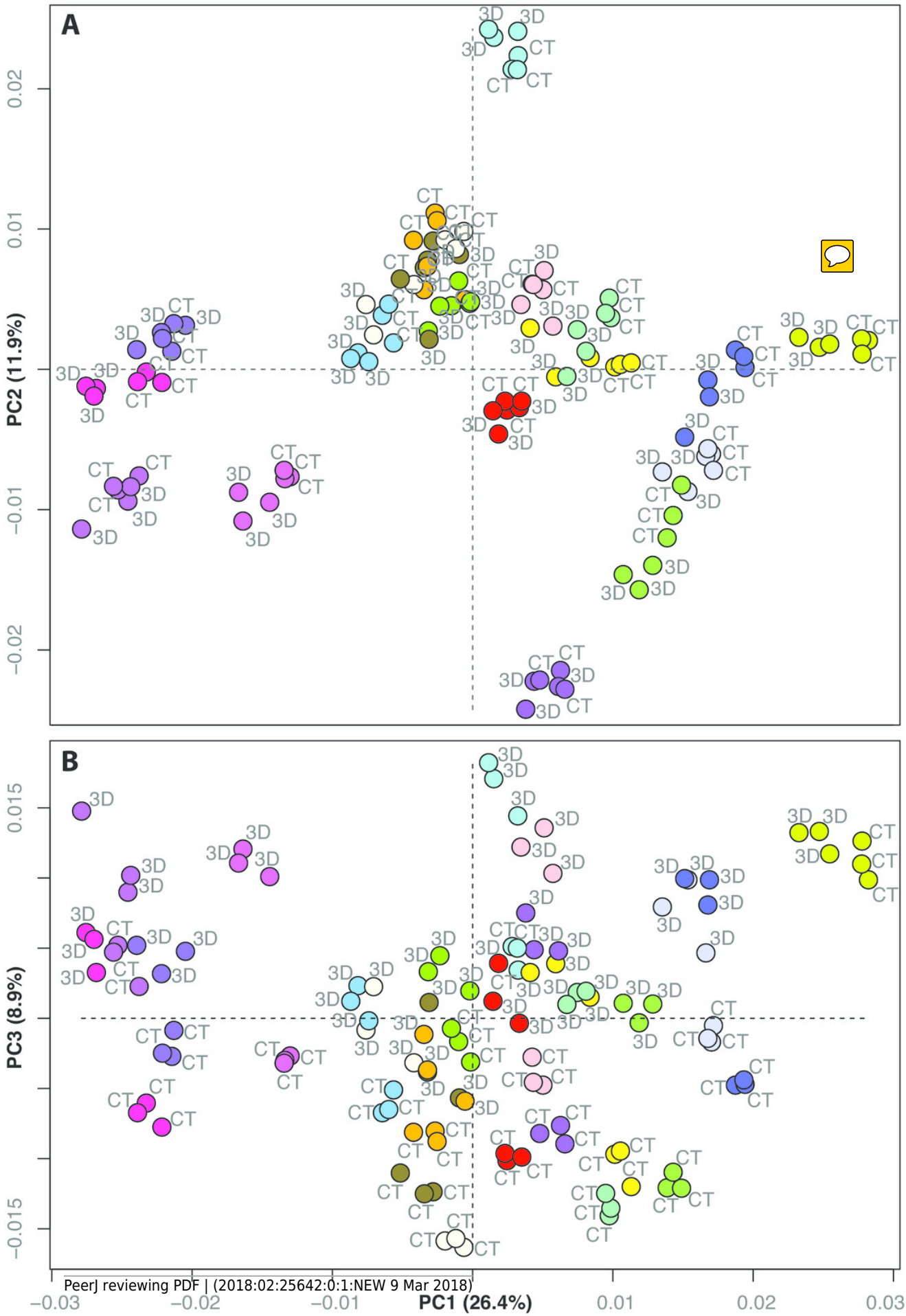

A
B
PC2 (11.9%)
PC3 (8.9%)
PC1 (26.4%)
0.02
0.01
−0.01
−0.02
0.015
−0.015
−0.03
−0.02
−0.01
0.01
0.02
0.03
3D
CT

Figure 5

3D warp-grids for the three most important principal components, showing minimum and maximum shapes for each PC.

The left hand cranium shows the minimum negative value for the PC and the right hand cranium shows the maximum positive value. (A) Positive values along PC1 (26.4% variance) correspond to a larger braincase relative to the rostrum. (B) Positive values along PC2 (11.9% variance) correspond to a wider frontal bone. (C) Positive values along PC3 (8.9% variance) correspond to a more dorsally-curved ventral surface.

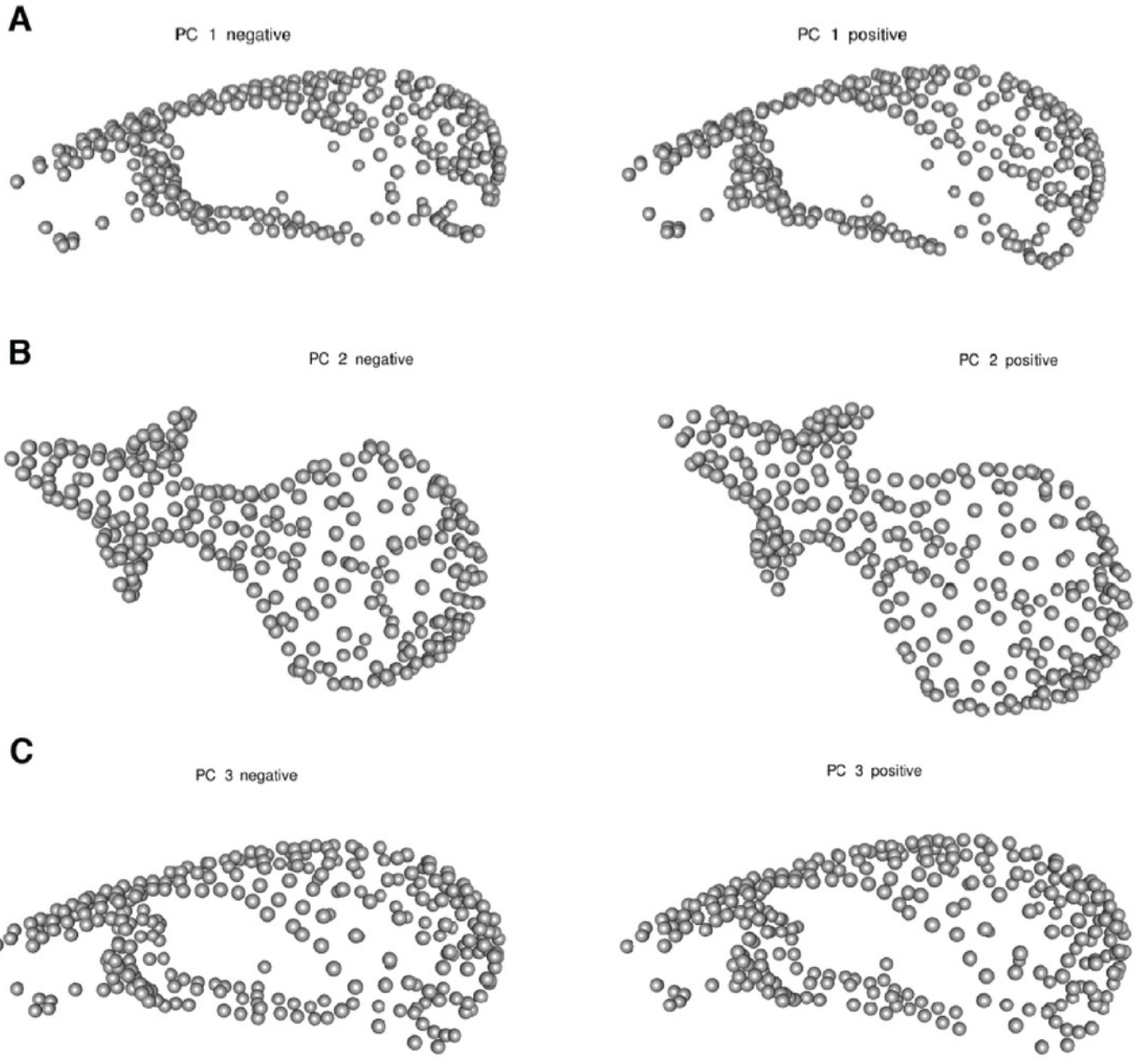


Figure 6

Morphological disparity -- as measured by shape variation among replicate scan triads -- by scanning device reflects operator error.

This box plot summarizes the morphological disparity (also known as the Procrustes variance) among the three replicates of an individual for each scan type. The mean Procrustes variance for 3D scans was $1.34x10^{-4}$ and $4.81x10^{-5}$ for µCT scans. This is a significant difference ($p<0.001$)

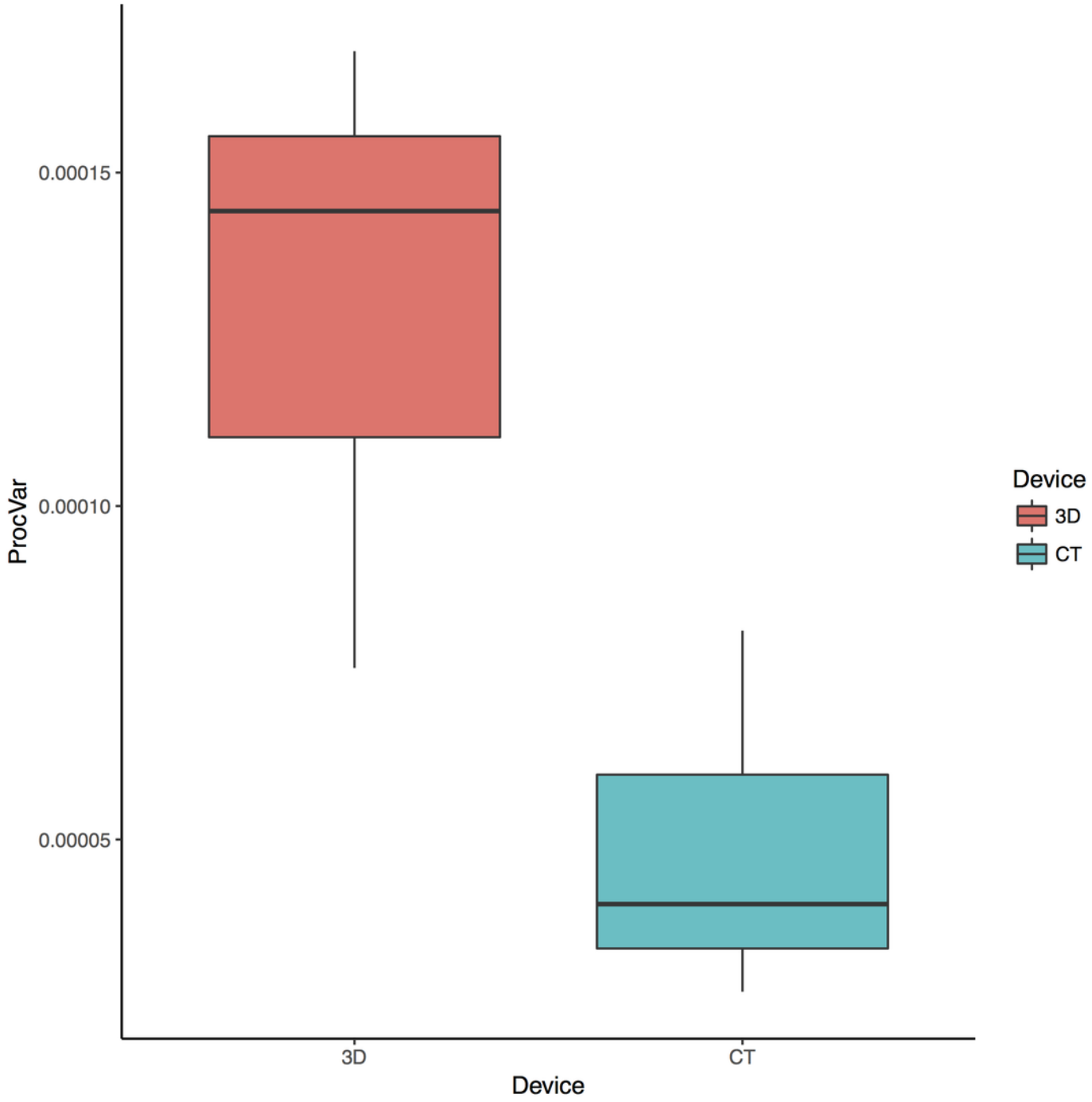
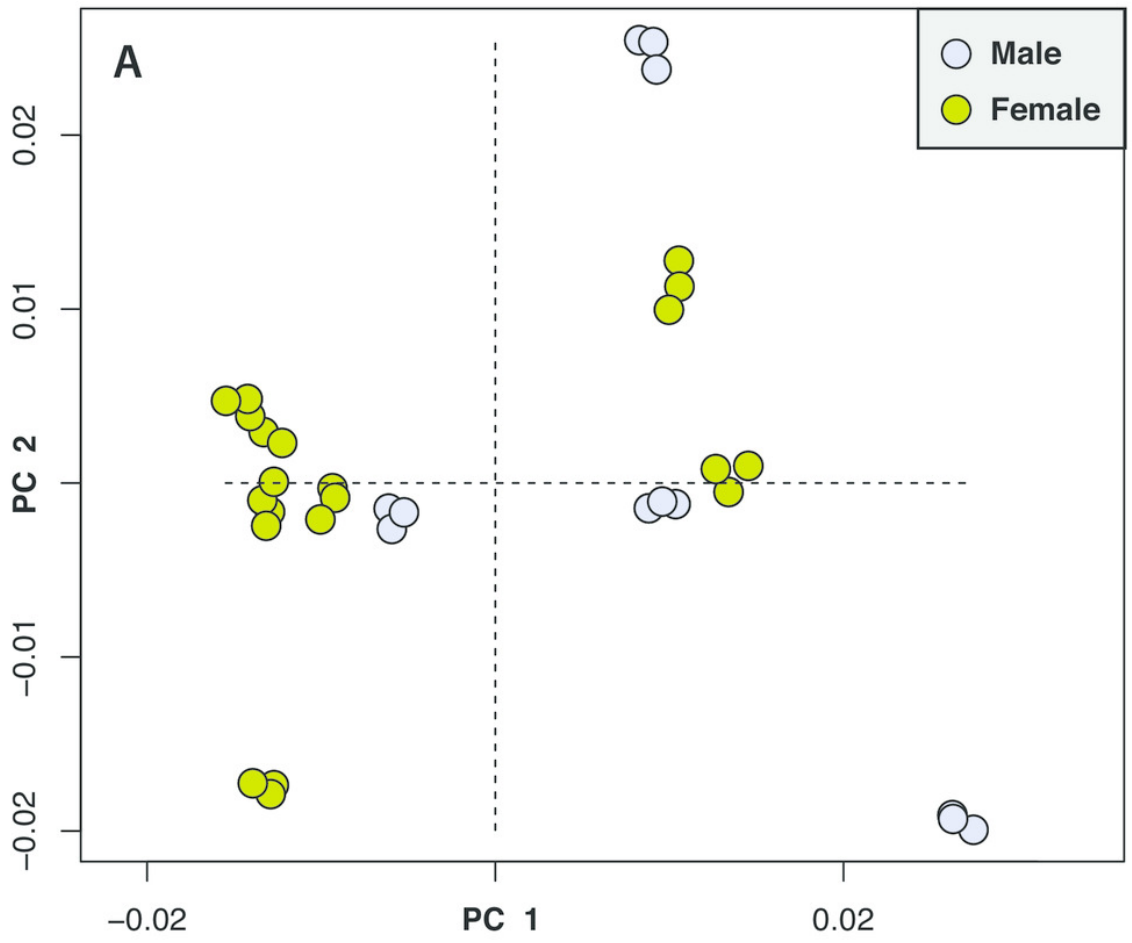

0.00015
0.00010
0.00005
ProcVar
3D
CT
Device
Device
3D
CT

Figure 7

Intra-specific variation as shown by PCAs of 3D and µCT scan datasets colored by sex.

PCA provides an exploratory visualization of shape variation between males and females in our subsample with sex information (n=11). Males (n=4) are plotted in light silver and females (n=7) are plotted in dark gold. Results from the cross-validation test can be found in Table 5.


A
Male
Female
0.02
0.01
−0.01
−0.02
PC 2
−0.02
PC 1
0.02

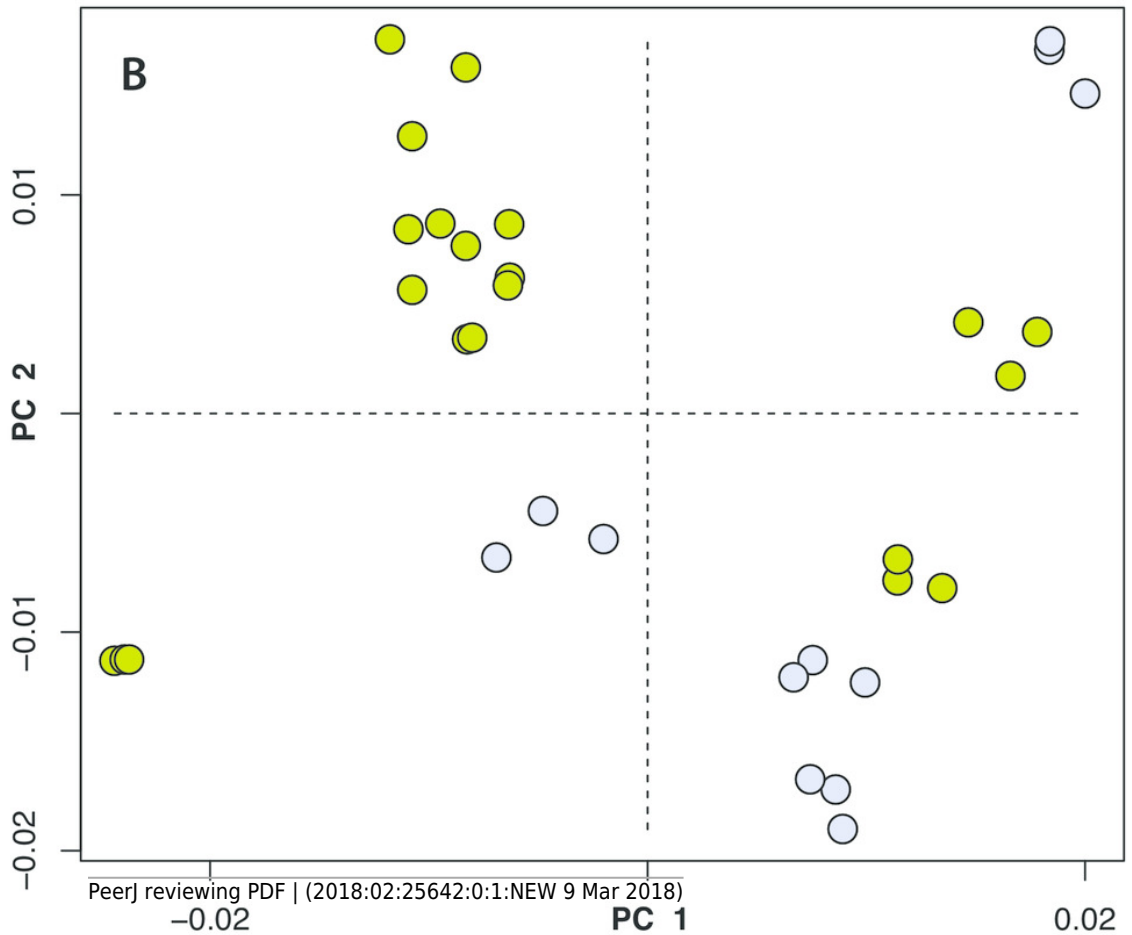

B
0.01
−0.01
−0.02
PC 2
−0.02
PC 1
0.02

Table 1(on next page)

General Procrustes ANOVA on sources of shape variation including asymmetry.

The %Var column of this Procrustes ANOVA demonstrates the relative contribution of each factor to overall variation. It is calculated from the sum of squares for each factor divided by the total sum of squares for all factors.

	Df	SS	MS	%Var	F	Pr(>F)
Individual	7740	0.06188221	7.9951E-06	47.4	11.12	<.0001
Side	400	0.0255547	6.38868E-05	19.6	88.89	<.0001
Ind * Side	7200	0.00517466	7.187E-07	4.0	0.55	1
Device	15770	0.02065404	1.3097E-06	15.8	4.79	<.0001
Res / Rep	63080	0.01723758	2.733E-07	13.2		

1

Table 2(on next page)

Procrustes ANOVA on the sources of shape variation using the symmetric component of shape.

The R-squared column of this Procrustes ANOVA demonstrates the relative contribution of each factor to overall variation. The shape variation of this dataset is visualized in Figures 4 and 5.

	Df	SS	MS	Rsq	F	Z	Pr (>F)
ind	18	0.062014315	0.00344524	0.73269356	25.31699532	21.2972812	0.001
ind: dev	19	0.01228211	0.00064643	0.14511204	4.75020269	23.624144	0.001
Resi-duals	76	0.010342389	0.000136084				
Total	113	0.084638816					

1

Table 3(on next page)

Comparison of operator error in 3D scan and µCT scan datasets using Procrustes ANOVAs and repeatability scores.

The repeatability score is a value that reflects the ease of digitizing in a repeated measure study design. It is calculated from the Procrustes ANOVA using formulas for the intra-class correlation coefficient. The Procrustes ANOVAs were found by subsetting the dataset by scan device and performing separate generalized Procrustes and bilateral symmetry alignments. (A) Results from the µCT-only dataset. (B) Results from the 3D-only dataset.

A								
	Df	**SS**	**MS**	**Rsq**	**F**	**Z**	**Pr(>F)**	**Repeatability**
µCT_ind	18	0.034310829	0.001906157	0.92599563	26.41573276	18.27750829	0.001	0.927
Residuals	38	0.002742077	7.22E-05					
Total	56	0.037052906						
B								
	Df	**SS**	**MS**	**Rsq**	**F**	**Z**	**Pr(>F)**	**Repeatability**
3D_ind	18	0.035295179	0.001960843	0.822025177	9.750741438	15.83823468	0.001	0.814
Residuals	38	0.00764168	0.000201097					
Total	56	0.042936859						

1

Table 4(on next page)

Symmetric Procrustes ANOVA with sex as a factor to assess relative contribution of intra-specific variation to overall shape variation.

This Procrustes ANOVA allows comparison of the relative contribution to total variation from sex and from scan device (R-squared column).

	df	SS	MS	Rsq	F	P
Ind	8600	0.03179244	3.6968E-06	0.6914	4.43	<.0001
Device	9460	0.00790042	8.351E-07	0.1718	5.03	<.0001
Sex/Res	37840	0.00628842	1.662E-07	0.1368		
Total	55900	0.04598128				

1

Table 5(on next page)

Between group PCA classification test to assess whether one scan device dataset performs better at identifying sexes based on shape.

This analysis averages shape among replicates, computes a between-group PCA separately for µCT and 3D datasets, and runs a cross-validation classification test. The results indicate whether one type of scan dataset is more successful at classifying males versus females based on the shape variation present in the dataset. It also returns a kappa statistic; a kappa value over 0.20 indicates “fair” agreement between the two datasets. Shape variation visualized by sex can be seen in Figure 7.

Cross-validated classification results in frequencies						
CT	**f**	**m**		**3D**	**f**	**m**
f	5	2		**f**	5	2
m	2	2		**m**	2	2
Cross-validated classification results in %						
CT	**f**	**m**		**3D**	**f**	**m**
f	71	29		**f**	71	29
m	50	50		**m**	50	50
Overall classification accuracy (%)						
CT	64					
3D	64					
Kappa statistic						
CT	0.214					
3D	0.214					

1