1 **Effects of different segmentation methods on geometric morphometric**

2 **data collection from primate skulls**

3

4 Tsuyoshi Ito

5

6 Department of Evolution and Phylogeny, Primate Research Institute, Kyoto

7 University, Inuyama, Aichi 484-8506, Japan

8 E-mail: ito.tsuyoshi.3a@kyoto-u.ac.jp

9

10 Running headline: Effects of segmentation in morphometrics

11

Abstract

12

13   1.   Increasing numbers of studies are analysing the shapes of objects using

14   geometric morphometrics with tomographic data, which are often segmented

15   and transformed to three-dimensional (3D) surface models before measurement.

16   The present study aimed to evaluate the effects of different image

17   segmentation methods on geometric morphometric data collection using

18   computed tomography data collected from non-human primate skulls.

19   2.   Three segmentation methods based on a visually-selected threshold, a

20   half-maximum height protocol and a gradient and watershed algorithm were

21   compared. For each method, the efficiency of surface reconstruction, the

22   accuracy of landmark placement and the level of variation in shape and size

23   compared with various levels of biological variation were evaluated.

24   3.   The visual-based method inflated the surface in high-density anatomical

25   regions, whereas the half-maximum height protocol resulted in large numbers

26   of artificial holes and erosion. However, the gradient-based method mitigated

27   these issues and generated the most efficient surface model. The segmentation

28   method used had a much smaller effect on shape and size variation than

29   interspecific and inter-individual differences. However, this effect was

30   statistically significant and not negligible when compared with intra-individual

31   (fluctuating asymmetric) variation.

32   4.   Although the gradient-based method is not widely used in geometric

33   morphometric analyses, it may be one of promising options for reconstructing

34   3D surfaces. When evaluating small variations, such as fluctuating asymmetry,

35   care should be taken around combining 3D data that were obtained using

36   different segmentation methods.

2

1. Introduction

Geometric morphometrics is a statistical method that uses Cartesian landmark coordinates to analyse the shapes of objects (Adams, Rohlf, & Slice, 2004, 2013; Slice, 2007; Mitteroecker & Gunz, 2009; Zelditch, Swiderski, & Sheets, 2012). This approach has the advantages of typically capturing a large number of shape variables and allowing changes in shape to be visualised (Mitteroecker & Gunz, 2009), which generally enables greater flexibility in data acquisition, more sensitive detection of shape variation and easier visual interpretation of the results compared with traditional morphometrics (Rohlf & Marcus, 1993; Parsons, Robinson, & Hrbek, 2003; Bernal, 2007; Maderbacher et al., 2008; van der Niet, Zollikofer, León, Johnson, & Linder, 2010; Breno, Leirs, & Van Dongen, 2011). Geometric morphometric data are appropriate for many different types of studies that seek to explain patterns of shape variation and the covariation of shape with other variables, making them useful for answering a wide variety of ecological and evolutionary questions, e.g. modularity, allometry, evolutionary process, ecological, geographical and phylogenetic diversities (Table S1). Consequently, increasing numbers of studies are using geometric morphometrics (Lawing & Polly, 2010), particularly in the fields of evolutionary biology (1,382), zoology (833), anthropology (674), ecology (531), genetics & heredity (379) and the others (3,377) [the numbers in parentheses indicate the number of literatures (1990–2018) retrieved from Web of Science using the search term 'geometric morphometric$'] (Fig. S1).

Three-dimensional (3D) landmark coordinates for geometric morphometric analyses are often captured using 3D surface models (Mitteroecker & Gunz, 2009; van der Niet et al., 2010; Adams & Otárola-Castillo, 2013). These models make it easier to obtain semi-landmarks on surfaces and thus to capture topography that lacks anatomical landmarks (Gunz, Mitteroecker, & Bookstein, 2005; Gunz & Mitteroecker, 2013) compared with the direct digitisation of an object using a contact-based 3D digitizing device such as MicroScribe (Immersion Corp., San Jose, USA). Furthermore, when sharing data, it is often preferable to deposit the original

3

3D data, such as surface models, rather than digitised coordinate data because the latter carries the risk of inter-observer error (Shearer et al., 2017). 3D surface models can be obtained with a 3D scanner (Friess, 2012), generated by photogrammetry (Evin et al., 2016; Muñoz-Muñoz, Quinto-Sánchez, & González-José, 2016) or extracted from tomographic data (two-dimensional image stacks) using procedures such as computed tomography (CT), micro-CT, magnetic resonance imaging (MRI), and micro-MRI (Vannier, Conroy, Marsh, & Knapp, 1985; Lorensen & Cline, 1987). However, 3D scanning and photogrammetry can capture only the surface of the object that is visible from the outside, while tomographic data can capture the entire body, including the internal structure (Conroy & Vannier, 1984), which is essential in some situations, such as where the shape of a precious fossil specimen in sediments or a rock matrix needs to be captured non-invasively. Thus, recent advances in information technology and enhancements in computer performance combined with the requirement for non-invasive measurements have led to 3D surface and tomographic data being increasingly recognised as important tools in biology (Cunningham, Rahman, Lautenschlager, Rayfield, & Donoghue, 2014; Baird & Taylor, 2017) (Fig. S1).

When using tomographic data, the surface is usually reconstructed once the region of interest (ROI) has been segmented from the remainder of the image. Several segmentation methods are available, each of which has its own advantages and disadvantages. Manual segmentation is still commonly used but is both labour-intensive and time-consuming and may carry the risk of human error (Stammberger, Eckstein, Michaelis, Englmeier, & Reiser, 1999). Threshold-based segmentation techniques are probably the most widely used, whereby a threshold is often determined by visual judgement (Heuzé et al., 2010; Toro-Ibacache, 2013; Noback & Harvati, 2015; Ito & Nishimura, 2016) or using the half-maximum height (HMH) protocol (Coleman & Colbert, 2007). While the visual-based method is probably the easiest to apply, its accuracy is unclear. By contrast, the HMH protocol calculates the threshold value as the mean of the maximum and minimum

grey values along a row of pixels that spans the boundary transition (Spoor, Zonneveld, & Macho, 1993), which provides accurate measurements (Coleman & Colbert, 2007). However, this approach often cannot efficiently segment the entire ROI if it exhibits a heterogeneous grey-value distribution (Rathnayaka, Sahama, Schuetz, & Schmutz, 2011), in which case multiple local thresholds can be calculated for a single ROI (Kubo, Kono, Saso, Mizushima, & Suwa, 2008), but this is a bit time-consuming. These issues seem to be solved by performing segmentation based on the grey-value gradient (edge detection) (Scherf & Tilgner, 2009; Rathnayaka et al., 2011). However, this requires some programming skills, as commonly used graphical user interface software often does not implement functions that can efficiently perform this. Since different segmentation methods may provide different results, care should be taken when using the reconstructed surfaces to obtain measurements.

As with traditional morphometrics, the validity of geometric morphometrics suffers from the effects of measurement error (Arnqvist, Mårtensson, & Hungaricae, 1998). Researchers sometimes evaluate the repeatability of measurements before undertaking the main analyses (Klingenberg & McIntyre, 1998; Willmore, Klingenberg, & Hallgrimsson, 2005; White & Searle, 2008; Viscosi & Cardini, 2011). Furthermore, systematic surveys have been conducted to examine intra-observer, inter-observer and inter-device (MicroScribe digitiser, 3D scanner, photogrammetry and/or CT) measurement errors and device-induced random errors (von Cramon-Taubadel, Frazier, & Lahr, 2007; Fruciano et al., 2017; Robinson & Terhune, 2017; Shearer et al., 2017; Marcy, Fruciano, Phillips, Mardon, & Weisbecker, 2018). These studies have demonstrated that although the inter-device error is much smaller than the inter-observer error (Shearer et al., 2017), it is still significant (Fruciano et al., 2017; Robinson & Terhune, 2017), with particularly large differences being observed between MicroScribe and the other devices (Robinson & Terhune, 2017) and low-cost 3D scanners tending to produce larger amounts of random error than high-resolution

127　micro-CT systems (Marcy et al., 2018). Thus, it is generally agreed that care should

128　be taken when combining data collected by different devices, particularly when

129　evaluating small amounts of biological variation, such as intra-individual and

130　intraspecific variations.

131　　　　　While great effort has been made to evaluate the effects of measurement

132　error on the collection of geometric morphometric data, little remains known about

133　the effects of different segmentation methods. A small number of papers have

134　reported on the effects of segmentation methods using a subset of samples prior to

135　the main analyses. One significant study is that of Toro-Ibacache (2013), which

136　showed that variation between the HMH- and visual-based segmentation methods

137　was much smaller than intraspecific variation in human skulls. By contrast, Gunz

138　et al (2012) reported that different threshold levels considerably affected surface

139　measurements of the bony labyrinth, while Ranthnayaka et al (2011) detected

140　significant differences in the thickness of the reconstructed long bone surface

141　between HMH-based and grey-value gradient-based (Canny-edge detection)

142　methods, although this was not specifically in relation to geometric morphometric

143　data collection. However, while it is clear that the threshold choice will have a

144　larger effect on the measurements when lower resolution images are used (i.e.

145　where there is a larger voxel size relative to the object size) (Hassan, Souza, Jacobs,

146　de Azambuja Berti, & van der Stelt, 2010; Gunz et al., 2012), it is not yet fully

147　understood how and to what extent different segmentation methods affect

148　measurements at a specific resolution. Therefore, to further expand our knowledge

149　in this area, the effects of various segmentation methods on geometric

150　morphometric data collection need to be systematically evaluated, alongside a

151　comparison with various levels of biological variation.

152　　　　　The aim of the present study was to evaluate the effects of three

153　segmentation methods (visual-, HMH- and grey-value gradient-based methods) on

154　geometric morphometric data collection from non-human primate skulls and to

155　compare inter-method variation in the shape data with interspecific,

156 inter-individual and intra-individual variations, as well as intra-observer error.
157 The gradient-based method was conducted by implementing the code of the
158 user-friendly cross-language programming interface SimpleITK (Lowekamp, Chen,
159 Ibáñez, & Blezek, 2013; Yaniv, Lowekamp, Johnson, & Beare, 2018) in the
160 open-source software Python programming language (Python Software Foundation,
161 https://www.python.org/) for easy sharing, testing and further improvements.

162 **2. Materials and Methods**

163 **2.1. Samples**

164 The dried crania of 19 adult males belonging to six species of the genus
165 *Macaca* were obtained from the Primate Research Institute of Kyoto University
166 (Inuyama, Japan) (Table 1). The specimens were scanned using the Asteion
167 Premium 4 helical CT scanner (Toshiba Medical Systems, Otawara, Japan) with a
168 slice thickness of 0.5 mm and serial CT images were reconstructed from the
169 original volumetric data using a pixel size ranging from 0.232 mm^2 to 0.348 mm^2
170 and an interslice interval of 0.2 mm.

171 **2.2. Image segmentation**

172 The images were processed using SimpleITK in Python unless otherwise
173 stated. The CT images were first resampled using linear interpolation to make
174 them cubic with a voxel size of 0.15 mm^3. The resampled images were then
175 denoised while preserving the edges using a curvature flow filter with a time step
176 of 0.01 and five iterations.

177 The resampled and denoised images were segmented using three
178 alternative methods (Fig. 1). The segmentation was not manually edited for any of
179 the methods.

180 The first method involved visually judging a global threshold for
181 segmentation (Fig. 1c). Using the Amira 6 software (FEI Visualisation Sciences
182 Group, Bordeaux, France), a global threshold was manually selected for each
183 specimen to obtain an optimum value where no scanning artefacts could be seen

184 and as many bony regions as possible were visible (e.g. Noback & Harvati, 2015).

185 This process was repeated twice for each specimen and the mean value was then

186 calculated and used in the following analyses [note: no specimen exhibited an error

187 of >10 Hounsfield units (HU)].

188   In the second method, a global threshold was calculated based on the

189 HMH protocol (Coleman & Colbert, 2007) (Fig. 1d). For each specimen, the HMH

190 was calculated from randomly drawn lines in randomly selected coronal slices

191 using the 'SCIKIT-IMAGE' package (van der Walt et al., 2014) in PYTHON. The global

192 threshold was then calculated as the mean HMH value of multiple lines. To target

193 only those lines that passed through the bone-to-air transition, any lines for which

194 the minimum HU was outside the range of −2,000 to −950 and the maximum HU

195 was outside the range of −800 to 3,000 were removed from the calculation. While

196 this approach does not necessarily guarantee that lines pass through bone-to-air

197 transition and may have underestimated the threshold compared with manually

198 drawing the lines, it is expected to be more stable where a larger number of lines

199 are used. To evaluate the repeatability of this method, this process was repeated 10

200 times for each specimen using 10, 100, 1,000 and 10,000 lines and the intraclass

201 correlation coefficient (ICC) was calculated using the 'ICC' package (Wolak,

202 Fairbairn, & Paulsen, 2012) in R software (R Developmental Core Team, 2019). In

203 the following analyses, the median value of 10 replicates with 10,000 lines was

204 taken as the global threshold for segmentation.

205   In the third method, segmentation was performed based on the image

206 gradient with watershed (GWS) algorithm (Withey & Koles, 2008; Aly, Bin Deris, &

207 Zaki, 2011) (Fig. 1e). Sobel filter was first applied to each image as this is

208 considered an efficient gradient detector (Senthilkumaran & Rajesh, 2009). The

209 gradient magnitude was then calculated, which functions as the topographic

210 surface in the watershed algorithm. Seeds for the watershed algorithm were

211 selected based on threshold-based segmentation. Seeds for bone were obtained by

212 selecting the bony region that was within an HU range of −500 to 10,000 and was

213   connected to a region with an HU > 0 and by eroding the selected region so that it

214   did not exceed the bone–air boundary. Seeds for air were obtained by selecting the

215   region that was within an HU range of −10,000 to −1,000. The watershed algorithm

216   was then used to segment the bony regions based on the gradient magnitude and

217   seeds. These fully automated procedures were carried out for all specimens using

218   the same parameters.

219   The accuracy of each segmentation method was evaluated by comparing

220   the results with a local HMH value as a reference. This evaluation was conducted

221   in two anatomical regions: the anterior zygomatic arch, which is usually hard and

222   has a high density and a high-HU value, and the posterior maxilla, which is often

223   thin and has a low density and a low-HU value. In each region, the segmentation

224   boundaries were detected on a line passing through the bone–air transition in a

225   coronal slice using the 'scikit-image' package in Python. The locations of the

226   segmentation boundaries of the three alternative models (the visual-, HMH- and

227   GWS-based models) were then compared with that of the reference model (local

228   HMH) along this line (Fig. 1b).

229   **2.3. 3D surface reconstruction and landmark-based analyses**

230   The segmented images were subjected to 3D surface reconstruction using

231   the 'generate surface' module of Amira 6 with the options compactify and

232   unconstrained smoothing (smoothing extent of five). The generated surfaces were

233   saved in the 'ply' format.

234   The ply-format data were first duplicated and linkable anonymised so that

235   an observer could take two measurements in a random order without knowing the

236   segmentation method or specimen ID. In total, 40 3D landmarks were obtained

237   from the duplicated and anonymised surface models by a single observer (TI) using

238   the Stratovan Checkpoint software (Stratovan Cor., Sacramento, CA, USA) (Table

239   S2; Fig. S2). If the observer made a different judgement about whether a landmark

240   was missing or not for the two replicates (which occurred in four instances), both

241   were treated as missing. The effect of segmentation method on the placement of

242 landmarks was examined by calculating the distances between homologous
243 landmarks among the three segmentation models using the average values for the
244 replicates.

245   Missing landmarks were estimated by mapping the weighted averages
246 from the complete dataset onto the specimen with missing values using the
247 'MORPHO' package (Schlager, 2017) in R. This was undertaken separately for the
248 three models. A generalised Procrustes analysis (GPA) was then performed to
249 register the landmark configurations using the 'GEOMORPH' package (Adams &
250 Otárola-Castillo, 2013) in R. The resulting GPA-registered shape data were
251 subjected to principal component analysis to visualise shape variations, and a
252 Procrustes analysis of variance (ANOVA) was performed to evaluate the effect of
253 segmentation method compared with individual, species, asymmetry and random
254 effects using the MORPHOJ software (Klingenberg, 2011).

255 **3. Results**

256 **3.1. 3D surface reconstruction**

257   The visual-based model was much more efficient in reconstructing the
258 bony regions than the HMH-based model but resulted in slightly larger and more
259 numerous artificial holes (i.e. holes that were generated in the surface models but
260 were not actually present) than the GWS-based model (Fig. 2). The visual-based
261 model agreed relatively well with the reference model at the posterior maxilla,
262 where the bony density was low (Fig. 3). However, the results were often up to 2
263 pixels thicker than for the reference model in the high-density anterior zygomatic
264 arch.

265   The HMH-based model resulted in many artificial holes, particularly at
266 the posterior maxilla, occipital and sphenoid bones (Fig. 2). This model showed
267 good agreement with the reference model in the anterior zygomatic arch region but
268 was often thinner than the reference model or did not lead to segmentation at the
269 posterior maxilla (Fig. 3). The use of 10 random lines was found to be insufficient to

10

270 efficiently replicate the results for segmenting skulls, likely due to the

271 heterogeneous distribution of HU values (Table 2; Fig. S3). Consequently, when

272 lines are randomly drawn, at least 100 or ideally thousands of lines should be used

273 to select a global threshold.

274   The GWS-based model resulted in fewer artificial holes than both the

275 visual- and HMH-based models, although the presence of artificial holes could not

276 be completely ruled out (Fig. 2). This model also consistently showed good

277 agreement with the reference model at both the anterior zygomatic arch and the

278 posterior maxilla (Fig. 3).

3.2. Landmarks and geometric morphometrics

280   There were one or two missing landmarks (RHI and/or HOR) in three

281 specimens for the visual-based model, 1–6 missing landmarks (RHI, PIF, PNS,

282 HOR, ZMI and/or PGP) in 17 specimens for the HMH-based model and one or two

283 missing landmarks (RHI and/or PIF) in two specimens for the GWS-based model. It

284 should be noted that HOR was missing in most (17 out of 19) specimens for the

285 HMH-based model.

286   The levels of inter-method and intra-observer error were evaluated for

287 each landmark. The inter-method error between the GWS- and visual-based models

288 was generally almost equal to or smaller than that between the GWS- and

289 HMH-based models and between the HMH- and visual-based models (Table S3; Fig.

290 4). Furthermore, the inter-method error between the GWS- and visual-based

291 models was not significantly larger than the intra-observer error (Table S3). By

292 contrast, the inter-method error between the HMH-based model and each of the

293 other two models exceeded the intra-observer error for some landmarks, including

294 PIF and MM1. The level of intra-observer error was not significantly different

295 among the three methods for any landmark except RHI (ANOVA, $P = 0.0349$) and

296 HOR ($P = 0.0499$) (Fig. S4).

297   Shape variations were assessed in the context of geometric morphometrics.

298 The first four principal components (PCs) accounted for 73.1% of the total shape

299 variance. The scatterplot of PC1 (37.7%) and PC2 (20.7%) reflected the level of

300 interspecific variation well (Fig. 5a), while that of PC3 (8.1%) and PC4 (6.7%)

301 explained inter-individual (intraspecific) variation (Fig. 5b). In both plots, the

302 inter-method and intra-observer errors appeared to be much smaller than the

303 interspecific and inter-individual variations. This was supported by the Procrustes

304 ANOVA, which showed that species and individual explained much more of the

305 variance than the method and residuals (Table 3), with the intra-individual

306 variance (fluctuating asymmetry) being nearly four-fold larger than the

307 inter-method variance. However, although the effect of method was small, it was

308 significant for both size and shape.

309 **4. Discussion**

310 **4.1. 3D surface reconstruction**

311 Visually judging a global threshold is probably the easiest and potentially

312 most commonly used method for image segmentation during 3D reconstruction

313 (Heuzé et al., 2010; Toro-Ibacache, 2013; Noback & Harvati, 2015; Ito & Nishimura,

314 2016). With this method, users generally select a global threshold where no

315 scanning artefacts can be seen and bony regions are as visible as possible (e.g.

316 Noback & Harvati, 2015). Artificial holes usually become more of a problem than

317 artefacts when observing scans of dry bones. Therefore, by its very nature, the

318 visual-based method is predicted to provide a lower threshold and therefore a more

319 inflated surface than is actually present on average. Supporting this, the present

320 study demonstrated that although the visual-based method accurately segmented

321 the low-density posterior maxilla, it inflated the surface of the high-density anterior

322 zygomatic arch. However, this inflation was only by 1–2 pixels, indicating that, if

323 using a high-resolution image, the visual-based method will give an accurate 3D

324 reconstruction to a certain extent.

325 The HMH-based method is also widely used in geometric morphometrics

326 due to its accuracy (Gröning, Fagan, & O'Higgins, 2011, 2012; Coquerelle et al.,

327 2013; Senck & Coquerelle, 2015; Stoessel, Gunz, David, & Spoor, 2016; Rivera &

328 Mirazón Lahr, 2017; Bauer, Benazzi, Darlas, & Harvati, 2018; Evteev, Anikin, &

329 Satanin, 2018). In the original paper describing this approach, the means of 10

330 HMH values along various randomly selected slices were used to obtain the global

331 threshold (Coleman & Colbert, 2007). Many studies have used this method or some

332 modified version of it, such as using a histogram of the grey values of a respective

333 volume rather than crossing lines to calculate the HMH (Senck, Bookstein, Benazzi,

334 Kastner, & Weber, 2015). However, although HMH is probably one of the most

335 accurate and best approaches for reconstructing a 3D surface when the object is

336 simple and has homogeneous grey-value distributions, it is often not suitable for a

337 complex object with heterogeneous grey-value distributions, such as a skull,

338 depending on the resolution of the image. Supporting this, the present study

339 demonstrated that the HMH-based model could not efficiently reconstruct the 3D

340 surface in the low-density posterior maxilla as it produced many artificial holes and

341 erosion. To address this, users must manually edit the segmentation or set multiple

342 local thresholds across an object, which is not only labour-intensive and

343 time-consuming but may also introduce human error or arbitrariness. Another

344 potential problem with the HMH protocol is the arbitrariness of selecting the lines

345 or regions that are used to calculate the HMH, which affects its repeatability. This

346 study suggested that at least 100 and ideally thousands of randomly drawn lines

347 should be used when selecting a global threshold for skulls. Otherwise, lines should

348 be drawn on pre-defined regions like as Evteev, Anikin, & Satanin (2018) instead of

349 random selection. Thus, it might be difficult to apply this method to a complex

350 object such as a skull unless to some extent high-resolution images are available.

351   The use of a gradient-based method appears to mitigate these issues

352 mentioned above. Rathnayaka et al (2011) demonstrated that the Cany filter-based

353 method generated 3D models of long bones with a significantly higher accuracy

354 than a visual-based method, and Scherf and Tilgner (2009) showed that the Ray

355 casting algorithm had a higher efficiency and accuracy for the 3D reconstruction of

356   the trabecular bone than the HMH protocol. Similarly, the present study

357   demonstrated that the watershed algorithm induced less artificial holes, erosion

358   and inflation than both the visual- and HMH-based models in the 3D reconstruction

359   of the cranium. While Rathnayaka et al (2011), Scherf and Tilgner (2009) and the

360   present study all used approaches that were based on image gradients (edges) in

361   the first step of the algorithm, they used different methods of ROI designation once

362   the gradient had been calculated – Rathnayaka et al (2011) extracted ROI outlines

363   by removing the branches of edges, filling gaps and implementing 3 × 3

364   neighborhood pixel operations, while Scherf and Tilgner (2009) cast rays along the

365   surface-normals, resulting in the area outside the ROI becoming bright and the

366   region inside the ROI becoming dark. While the application of these methods

367   requires some script writing, this brings the added benefit of flexibility and

368   automaticy, and both MATLAB (Mathworks, MA, USA) and Python, which were

369   used in Rathnayaka et al (2011) and the present study, respectively, are

370   user-friendly languages. Thus, the gradient-based method may represent one of

371   promising options for the 3D surface reconstruction of skulls, athough it is not yet

372   widely used in the field of geometric morphometrics (but see Le Cabec, Kupczik,

373   Gunz, Braga, & Hublin, 2012; Le Cabec, Gunz, Kupczik, Braga, & Hublin, 2013;

374   Navarro & Maga, 2016; Hublin et al., 2017; Pan et al., 2017).

375   Segmenting images with lower contrast and/or more noise (e.g. when

376   differentiating between soft tissues or extracting fossils from surrounding

377   sediments/rocks) is more challenging. However, the application of automated

378   operations, such as gradient (edge) detection, morphological operations and

379   watershed algorithms, will probably eliminate or at least reduce the labor

380   associated with manual processing. Other than these, there are also many kinds of

381   segmentation methods (Pham, Xu, & Prince, 2000), some of which can potentially

382   efficiently handle low contrast and noisy images; for example, when segmenting the

383   same region across many individuals wihin a population, the atlas-based algorithm,

384   which is based on volume registration, is likely useful (Cabezas, Oliver, Lladó,

385 Freixenet, & Cuadra, 2011). Segmentation methods should be considered
386 depending on the image properties, the number of specimens and the required
387 accuracy for downstream analyses.

4.2. Landmarks and geometric morphometrics

389   Insufficient segmentation induces artificial holes and erosion and thus
390 increases the number of missing landmarks. The present study showed that the
391 HMH-based method could not efficiently reconstruct low-density regions, such as
392 those around HOR, so a global threshold based on this method is not necessarily
393 recommended for the 3D reconstruction of skulls unless high-resolution images are
394 available.

395   Analysis of the amount of inter-method error for each of the 40 landmarks
396 showed that the positioning of the landmarks was most similar between the GWS-
397 and visual-based methods, suggesting that the HMH-based method leads to
398 erroneous positioning. In particular, the positions of PIF and MM1 were very
399 different between the HMH-based model and the other two models, and these
400 differences significantly exceeded the intra-observer error. In addition, the
401 inter-method error between the HMH- and GWS-based methods also significantly
402 exceeded the intra-observer error for PRS and RHI. This indicates that landmarks
403 that are located on the tip (e.g. PIF, PRS, RHI) or in low-density regions (e.g. MM1)
404 are more susceptible to improper thresholding.

405   Significant differences in the level of intra-observer error among methods
406 were only observed for RHI and HOR. It is notable that the intra-observer error for
407 RHI was high with the GWS-based method, although the reason for this remains
408 unclear. Since the sample size of HOR was small for the HMH-based method, the
409 result for this landmark should be interpreted with caution. For most other
410 landmarks, the method that was used did not affect the likelihood of random error
411 occurring, which suggests that repeatability is good assuming only one
412 segmentation method is used.

413   Shape was assessed at various levels of biological variation. The image

segmentation method that was used had a much smaller effect on the variation in shape than the species or individual, supporting the findings of a previous evaluation of human skull variations (Toro-Ibacache, 2013). However, method still had a significant effect. Furthermore, the effect of method cannot be considered negligible when compared with the level of intra-individual variation (fluctuating asymmetry). Therefore, when evaluating small biological variations, such as fluctuating asymmetry, care should be taken if combining multiple sources of 3D data that were obtained using different segmentation methods.

There is growing recognition of the importance of open data and data sharing in terms of the verifiability of results, the reproducibility of studies and the reuse of data (Rowe & Frank, 2011; Wicherts & Bakker, 2012; Davies et al., 2017; Lowndes et al., 2017; Culina et al., 2018). Data sharing is particularly important for morphological studies, as it not only saves financial and labour resources but also helps protect specimens from damage by reducing the amount of repeated handling and measurement by different researchers (Davies et al., 2017). However, it is also true that data sharing has been considered to require some cautions, because measurement error is inevitable (Fruciano et al., 2017; Robinson & Terhune, 2017; Shearer et al., 2017; this study). Researchers willing to use shared 3D surface models which were reconstructed from tomographic data should consider the properties of original tomographic images, segmentation methods, and their potential impact on inference. The findings of the present study will form the basis for such consideration. Ideally, original image stacks (with sufficient header information) rather than (or in addition to) 3D models or landmarks are encouraged to be shared (see also Davies et al., 2017).

4.3. Limitations and future directions

The present study had several limitations and leaves some challenges for future studies. The effects of different segmentation methods on geometric morphometric data collection were evaluated using images obtained from a single helical CT scanner with similar parameter settings. However, the efficiency of 3D

443 surface reconstruction is greatly affected by the resolution and quality of the
444 images (Hassan et al., 2010), so higher resolution images are likely to generate a
445 better reconstruction regardless of which segmentation method is used
446 (Palacio-Mancheno, Larriera, Doty, Cardoso, & Fritton, 2014), while extremely low
447 resolution images (e.g. medical CT images of the mammalian bony labyrinth) will
448 be more susceptible to threshold selection (Gunz et al., 2012). At such an extremely
449 low resolution, the difference between the HMH- and visual-based methods may
450 not be negligible even at the inter-individual level, and it is also unclear how the
451 gradient-based method would perform. Image filtering will also affect the
452 segmentation, with too little smoothing likely retaining noise and thus causing
453 artefacts and too much smoothing readily producing artificial holes. Therefore,
454 these points should be kept in mind when applying the findings of the present
455 study to different kinds of images. GWS-based method that was implemented in
456 this study could not fully remove the artificial holes. This may have been partly due
457 to the seeds not being correctly or sufficiently assigned or the gradient not being
458 efficiently detected in low-intensity regions. It is expected that future research will
459 further improve the segmentation method and evaluate the validity of combining
460 images of various resolutions, which are processed with various filters.

461 **Data accessibility**
462   All of the CT data in dicom format are available at the Digital Morphology
463 Museum, Kyoto University (dmm.pri.kyoto-u.ac.jp/dmm/WebGallery/): The PRICT
464 Nos. are 325, 1257, 357, 322, 900, 1243, 920, 342, 989, 1001, 326, 1010, 1007, 387,
465 1005, 1245, 903, 323 and 1239. The other datasets and code used in this study are
466 available at Dryad Digital Repository (https://doi.org/10.5061/dryad.qk58nt2).

467 **Competing interests**
468   We declare we have no competing interests.

References

482
483   Adams, D. C., & Otárola-Castillo, E. (2013). geomorph: an R package for the
484   collection and analysis of geometric morphometric shape data. *Methods in*
485   *Ecology and Evolution*, *4*(4), 393–399. doi:10.1111/2041-210X.12035
486   Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten
487   years of progress following the 'revolution'. *Italian Journal of Zoology*, *71*(1), 5–
488   16. doi:10.1080/11250000409356545
489   Adams, D. C., Rohlf, F. J., & Slice, D. E. (2013). A field comes of age: Geometric
490   morphometrics in the 21st century. *Hystrix, the Italian Journal of Mammalogy*,
491   *24*(1), 7–14. doi:10.4404/hystrix-24.1-6283
492   Aly, A. A., Bin Deris, S., & Zaki, N. (2011). Research review for digital image
493   segmentation techniques. *International Journal of Computer Science and*
494   *Information Technology*, *3*(5), 99–106. doi:10.5121/ijcsit.2011.3509
495   Arnqvist, G., Mårtensson, T., & Hungaricae, A. S. (1998). Measurement error in
496   geometric morphometrics: empirical strategies to assess and reduce its impact
497   on measures of shape. *Acta Zoologica Academiae Scientiarum Hungaricae*,
498   *44*(1–2), 73–96.
499   Baird, E., & Taylor, G. (2017). X-ray micro computed-tomography. *Current Biology*,
500   *27*(8), R289–R291. doi:10.1016/j.cub.2017.01.066
501   Bauer, C. C., Benazzi, S., Darlas, A., & Harvati, K. (2018). Geometric morphometric
502   analysis and internal structure measurements of the Neanderthal lower fourth
503   premolars from Kalamakia, Greece. *Quaternary International*, *497*, 14–21.
504   doi:10.1016/j.quaint.2018.01.035
505   Bernal, V. (2007). Size and shape analysis of human molars: Comparing traditional
506   and geometric morphometric techniques. *HOMO- Journal of Comparative*
507   *Human Biology*, *58*(4), 279–296. doi:10.1016/j.jchb.2006.11.003

508   Breno, M., Leirs, H., & Van Dongen, S. (2011). Traditional and geometric
509   morphometrics for studying skull morphology during growth in Mastomys
510   natalensis (Rodentia: Muridae). *Journal of Mammalogy*, *92*(6), 1395–1406.
511   doi:10.1644/10-mamm-a-331.1
512   Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., & Cuadra, M. B. (2011). A review of
513   atlas-based segmentation for magnetic resonance brain images. *Computer*
514   *Methods and Programs in Biomedicine*, *104*(3), e158–e177.
515   doi:10.1016/j.cmpb.2011.07.015
516   Coleman, M. N., & Colbert, M. W. (2007). Technical note: CT thresholding protocols
517   for taking measurements on three-dimensional models. *American Journal of*
518   *Physical Anthropology*, *133*(1), 723–725.
519   Conroy, G. C., & Vannier, M. W. (1984). Noninvasive three-dimensional computer
520   imaging of matrix-filled fossil skulls by high-resolution computed tomography.
521   *Science*, *226*(4673), 456–458. doi:10.1126/science.226.4673.456
522   Coquerelle, M., Prados-Frutos, J. C., Benazzi, S., Bookstein, F. L., Senck, S.,
523   Mitteroecker, P., & Weber, G. W. (2013). Infant growth patterns of the
524   mandible in modern humans: a closer exploration of the developmental
525   interactions between the symphyseal bone, the teeth, and the suprahyoid and
526   tongue muscle insertion sites. *Journal of Anatomy*, *222*(2), 178–192.
527   doi:10.1111/joa.12008
528   Culina, A., Baglioni, M., Crowther, T. W., Visser, M. E., Woutersen-Windhouwer, S.,
529   & Manghi, P. (2018). Navigating the unfolding open data landscape in ecology
530   and evolution. *Nature Ecology and Evolution*, *2*(3), 420–426.
531   doi:10.1038/s41559-017-0458-2
532   Cunningham, J. A., Rahman, I. A., Lautenschlager, S., Rayfield, E. J., & Donoghue,
533   P. C. J. (2014). A virtual world of paleontology. *Trends in Ecology and*
534   *Evolution*, *29*(6), 347–357. doi:10.1016/j.tree.2014.04.004
535   Davies, T. G., Rahman, I. A., Lautenschlager, S., Cunningham, J. A., Asher, R. J.,
536   Barrett, P. M., … Donoghue, P. C. J. (2017). Open data and digital morphology.
537   *Proceedings of the Royal Society B: Biological Sciences*, *284*(1852), 20170194.
538   doi:10.1098/rspb.2017.0194
539   Evin, A., Souter, T., Hulme-Beaman, A., Ameen, C., Allen, R., Viacava, P., …
540   Dobney, K. (2016). The use of close-range photogrammetry in zooarchaeology:
541   Creating accurate 3D models of wolf crania to study dog domestication.
542   *Journal of Archaeological Science: Reports*, *9*, 87–93.
543   doi:10.1016/J.JASREP.2016.06.028
544   Evteev, A., Anikin, A., & Satanin, L. (2018). Midfacial growth patterns in males
545   from newborn to 5 years old based on computed tomography. *American Journal*
546   *of Human Biology*, *30*(4), e23132.
547   Friess, M. (2012). Scratching the surface? The use of surface scanning in physical
548   and paleoanthropology. *Journal of Anthropological Sciences*, *90*, 7–31.
549   doi:10.4436/jass.90004
550   Fruciano, C., Celik, M. A., Butler, K., Dooley, T., Weisbecker, V., & Phillips, M. J.
551   (2017). Sharing is caring? Measurement error and the issues arising from
552   combining 3D morphometric datasets. *Ecology and Evolution*, *7*(17), 7034–
553   7046. doi:10.1002/ece3.3256
554   Gröning, F., Fagan, M. J., & O'Higgins, P. (2011). The effects of the periodontal
555   ligament on mandibular stiffness: a study combining finite element analysis
556   and geometric morphometrics. *Journal of Biomechanics*, *44*(7), 1304–1312.
557   doi:10.1016/j.jbiomech.2011.01.008
558   Gröning, F., Fagan, M., & O'Higgins, P. (2012). Modeling the human mandible
559   under masticatory loads: Which input variables are important? *Anatomical*
560   *Record*, *295*(5), 853–863. doi:10.1002/ar.22455
561   Gunz, P., & Mitteroecker, P. (2013). Semilandmarks: A method for quantifying

562 curves and surfaces. *Hystrix, the Italian Journal of Mammalogy*, *24*(1), 103–
563 109. doi:10.4404/hystrix-24.1-6292
564 Gunz, P., Mitteroecker, P., & Bookstein, F. L. (2005). Semilandmarks in three
565 dimensions. In D. E. Slice (Ed.), *Modern morphometrics in physical*
566 *anthropology* (pp. 73–98). New York: Kluwer Academic Publishers.
567 Gunz, P., Ramsier, M., Kuhrig, M., Hublin, J.-J., & Spoor, F. (2012). The
568 mammalian bony labyrinth reconsidered, introducing a comprehensive
569 geometric morphometric approach. *Journal of Anatomy*, *220*(6), 529–543.
570 doi:10.1111/j.1469-7580.2012.01493.x
571 Hassan, B., Souza, P. C., Jacobs, R., de Azambuja Berti, S., & van der Stelt, P.
572 (2010). Influence of scanning and reconstruction parameters on quality of
573 three-dimensional surface models of the dental arches from cone beam
574 computed tomography. *Clinical Oral Investigations*, *14*(3), 303–310.
575 doi:10.1007/s00784-009-0291-3
576 Heuzé, Y., Boyadjiev, S. A., Marsh, J. L., Kane, A. A., Cherkez, E., Boggan, J. E., &
577 Richtsmeier, J. T. (2010). New insights into the relationship between suture
578 closure and craniofacial dysmorphology in sagittal nonsyndromic
579 craniosynostosis. *Journal of Anatomy*, *217*(2), 85–96.
580 doi:10.1111/j.1469-7580.2010.01258.x
581 Hublin, J. J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M.
582 M., … Gunz, P. (2017). New fossils from Jebel Irhoud, Morocco and the
583 pan-African origin of *Homo sapiens*. *Nature*, *546*(7657), 289–292.
584 doi:10.1038/nature22336
585 Ito, T., & Nishimura, T. D. (2016). Enigmatic diversity of the maxillary sinus in
586 macaques and its possible role as a spatial compromise in craniofacial
587 modifications. *Evolutionary Biology*, *43*(3), 1–13.
588 doi:10.1007/s11692-016-9369-4
589 Klingenberg, C. P. (2011). MorphoJ: an integrated software package for geometric
590 morphometrics. *Molecular Ecology Resources*, *11*(2), 353–357.
591 Klingenberg, C. P., & McIntyre, G. S. (1998). Geometric morphometrics of
592 developmental instability: analyzing patterns of fluctuating asymmetry with
593 Procrustes methods. *Evolution*, *52*(5), 1363–1375.
594 Kubo, D., Kono, R. T., Saso, A., Mizushima, S., & Suwa, G. (2008). Accuracy and
595 precision of CT-based endocranial capacity estimations: a comparison with the
596 conventional millet seed method and application to the Minatogawa 1 skull.
597 *Anthropological Science*, *116*(1), 77–85. doi:10.1537/ase.070502
598 Lawing, A. M., & Polly, P. D. (2010). Geometric morphometrics: recent applications
599 to the study of evolution and development. *Journal of Zoology*, *280*(1), 1–7.
600 doi:10.1111/j.1469-7998.2009.00620.x
601 Le Cabec, A., Gunz, P., Kupczik, K., Braga, J., & Hublin, J. J. (2013). Anterior tooth
602 root morphology and size in Neanderthals: Taxonomic and functional
603 implications. *Journal of Human Evolution*, *64*(3), 169–193.
604 doi:10.1016/j.jhevol.2012.08.011
605 Le Cabec, A., Kupczik, K., Gunz, P., Braga, J., & Hublin, J. J. (2012). Long anterior
606 mandibular tooth roots in Neanderthals are not the result of their large jaws.
607 *Journal of Human Evolution*, *63*(5), 667–681. doi:10.1016/j.jhevol.2012.07.003
608 Lorensen, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D
609 surface construction algorithm. *ACM Siggraph Computer Graphics*, *21*(4), 163–
610 169. doi:10.1145/37402.37422
611 Lowekamp, B. C., Chen, D. T., Ibáñez, L., & Blezek, D. (2013). The design of
612 SimpleITK. *Frontiers in Neuroinformatics*, *7*, 45. doi:10.3389/fninf.2013.00045
613 Lowndes, J. S. S., Best, B. D., Scarborough, C., Afflerbach, J. C., Frazier, M. R.,
614 O'Hara, C. C., … Halpern, B. S. (2017). Our path to better science in less time
615 using open data science tools. *Nature Ecology and Evolution*, *1*(6), 0160.

616   doi:10.1038/s41559-017-0160

617 Maderbacher, M., Bauer, C., Herler, J., Postl, L., Makasa, L., & Sturmbauer, C.
618   (2008). Assessment of traditional versus geometric morphometrics for
619   discriminating populations of the *Tropheus moorii* species complex (Teleostei:
620   Cichlidae), a Lake Tanganyika model for allopatric speciation. *Journal of*
621   *Zoological Systematics and Evolutionary Research*, *46*(2), 153–161.
622   doi:10.1111/j.1439-0469.2007.00447.x

623 Marcy, A. E., Fruciano, C., Phillips, M. J., Mardon, K., & Weisbecker, V. (2018). Low
624   resolution scans can provide a sufficiently accurate, cost- and time-effective
625   alternative to high resolution scans for 3D shape analyses. *PeerJ*, *6*, e5032.
626   doi:10.7717/peerj.5032

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

629 Muñoz-Muñoz, F., Quinto-Sánchez, M., & González-José, R. (2016).
630   Photogrammetry: a useful tool for three-dimensional morphometric analysis of
631   small mammals. *Journal of Zoological Systematics and Evolutionary Research*,
632   *54*(4), 318–325. doi:10.1111/jzs.12137

633 Navarro, N., & Maga, A. M. (2016). Does 3D phenotyping yield substantial insights
634   in the genetics of the mouse mandible shape? *G3: Genes|Genomes|Genetics*,
635   *6*(5), 1153–1163. doi:10.1534/g3.115.024372

636 Noback, M. L., & Harvati, K. (2015). The contribution of subsistence to global
637   human cranial variation. *Journal of Human Evolution*, *80*, 34–50.
638   doi:10.1016/j.jhevol.2014.11.005

639 Palacio-Mancheno, P. E., Larriera, A. I., Doty, S. B., Cardoso, L., & Fritton, S. P.
640   (2014). 3D assessment of cortical bone porosity and tissue mineral density
641   using high-resolution μCT: Effects of resolution and threshold method. *Journal*
642   *of Bone and Mineral Research*, *29*(1), 142–150. doi:10.1002/jbmr.2012

643 Pan, L., Thackeray, J. F., Dumoncel, J., Zanolli, C., Oettlé, A., de Beer, F., … Braga,
644   J. (2017). Intra-individual metameric variation expressed at the
645   enamel-dentine junction of lower post-canine dentition of South African fossil
646   hominins and modern humans. *American Journal of Physical Anthropology*,
647   *163*(4), 806–815. doi:10.1002/ajpa.23240

648 Parsons, K. J., Robinson, B. W., & Hrbek, T. (2003). Getting into shape: An
649   empirical comparison of traditional truss-based morphometric methods with a
650   newer geometric method applied to New World cichlids. *Environmental Biology*
651   *of Fishes*, *67*(4), 417–431. doi:10.1023/a:1025895317253

652 Pham, D. L., Xu, C., & Prince, J. L. (2000). Current methods in medical image
653   segmentation. *Annual Review of Biomedical Engineering*, *2*(1), 315–337.
654   doi:10.1146/annurev.bioeng.2.1.315

655 R Developmental Core Team. (2019). R: A language and environment for statistical
656   computing. *R Foundation for Statistical Computing*. Vienna.

657 Rathnayaka, K., Sahama, T., Schuetz, M. A., & Schmutz, B. (2011). Effects of CT
658   image segmentation methods on the accuracy of long bone 3D reconstructions.
659   *Medical Engineering & Physics*, *33*(2), 226–233.
660   doi:10.1016/j.medengphy.2010.10.002

661 Rivera, F., & Mirazón Lahr, M. (2017). New evidence suggesting a dissociated
662   etiology for cribra orbitalia and porotic hyperostosis. *American Journal of*
663   *Physical Anthropology*, *164*(1), 76–96. doi:10.1002/ajpa.23258

664 Robinson, C., & Terhune, C. E. (2017). Error in geometric morphometric data
665   collection: Combining data from multiple sources. *American Journal of*
666   *Physical Anthropology*, *164*(1), 62–75. doi:10.1002/AJPA.23257

667 Rohlf, F. J., & Marcus, L. F. (1993). A revolution in morphometrics. *Trends in*
668   *Ecology and Evolution.*, *8*(4), 129–132.
669   https ://doi.org/10.1016/0169-5347(93)90024-j

670 Rowe, T., & Frank, L. R. (2011). The disappearing third dimension. *Science,*
671   *331*(6018), 712–714. doi:10.1126/science.1202828
672 Scherf, H., & Tilgner, R. (2009). A new high-resolution computed tomography (CT)
673   segmentation method for trabecular bone architectural analysis. *American*
674   *Journal of Physical Anthropology, 140*(1), 39–51. doi:10.1002/ajpa.21033
675 Schlager, S. (2017). Morpho and Rvcg - Shape Analysis in R. In G. Zheng, S. Li, & G.
676   Szekely (Eds.), *Statistical Shape and Deformation Analysis: Methods,*
677   *Implementation and Applications* (pp. 217–256). San Diego: Academic Press.
678 Senck, S., Bookstein, F. L., Benazzi, S., Kastner, J., & Weber, G. W. (2015). Virtual
679   reconstruction of modern and fossil hominoid crania: Consequences of
680   reference sample choice. *Anatomical Record, 298*(5), 827–841.
681   doi:10.1002/ar.23104
682 Senck, S., & Coquerelle, M. (2015). Morphological integration and variation in facial
683   orientation in *Pongo pygmaeus pygmaeus*: A geometric morphometric approach
684   via partial least squares. *International Journal of Primatology, 36*(3), 489–512.
685   doi:10.1007/s10764-015-9837-1
686 Senthilkumaran, N., & Rajesh, R. (2009). Image segmentation - A survey of soft
687   computing approaches. *International Journal of Recent Trends in Engineering,*
688   *1*(2), 250–254. doi:10.1109/ARTCom.2009.219
689 Shearer, B. M., Cooke, S. B., Halenar, L. B., Reber, S. L., Plummer, J. E., Delson, E.,
690   & Tallman, M. (2017). Evaluating causes of error in landmark-based data
691   collection using scanners. *PLoS ONE, 12*(11).
692   doi:10.1371/journal.pone.0187452
693 Slice, D. E. (2007). Geometric morphometrics. *Annual Review of Anthropology, 36,*
694   261–281. doi:10.1016/B978-0-12-385189-5.00012-1
695 Spoor, C. F., Zonneveld, F. W., & Macho, G. a. (1993). Linear measurements of
696   cortical bone and dental enamel by computed tomography: applications and
697   problems. *American Journal of Physical Anthropology, 91*(4), 469–484.
698   doi:10.1002/ajpa.1330910405
699 Stammberger, T., Eckstein, F., Michaelis, M., Englmeier, K. H., & Reiser, M. (1999).
700   Interobserver reproducibility of quantitative cartilage measurements:
701   Comparison of B-spline snakes and manual segmentation. *Magnetic Resonance*
702   *Imaging, 17*(7), 1033–1042. doi:10.1016/S0730-725X(99)00040-5
703 Stoessel, A., Gunz, P., David, R., & Spoor, F. (2016). Comparative anatomy of the
704   middle ear ossicles of extant hominids – Introducing a geometric morphometric
705   protocol. *Journal of Human Evolution, 91*, 1–25.
706   doi:10.1016/J.JHEVOL.2015.10.013
707 Toro-Ibacache, M. V. (2013). *A finite element study of the human cranium ; The*
708   *impact of morphological variation on biting performance.* Kingston upon Hull,
709   UK: Hull York Medical School, The University of Hull and University of York.
710 van der Niet, T., Zollikofer, C. P. E., León, M. S. P. de, Johnson, S. D., & Linder, H.
711   P. (2010). Three-dimensional geometric morphometrics for studying floral
712   shape variation. *Trends in Plant Science, 15*(8), 423–426.
713   doi:10.1016/j.tplants.2010.05.005
714 van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D.,
715   Yager, N., … Yu, T. (2014). scikit-image: image processing in Python. *PeerJ, 2,*
716   e453. doi:10.7717/peerj.453
717 Vannier, M. W., Conroy, G. C., Marsh, J. L., & Knapp, R. H. (1985). Three-
718   dimensional cranial surface reconstructions using high-resolution computed
719   tomography. *American Journal of Physical Anthropology, 67*(4), 299–311.
720   doi:10.1002/ajpa.1330670402
721 Viscosi, V., & Cardini, A. (2011). Leaf morphology, taxonomy and geometric
722   morphometrics: A simplified protocol for beginners. *PLoS ONE, 6*(10).
723   doi:e25630 10.1371/journal.pone.0025630

724 von Cramon-Taubadel, N., Frazier, B. C., & Lahr, M. M. (2007). The problem of
725   assessing landmark error in geometric morphometrics: theory, methods, and
726   modifications. *American Journal of Physical Anthropology*, *134*(1), 24–35.
727 White, T. A., & Searle, J. B. (2008). Mandible asymmetry and genetic diversity in
728   island populations of the common shrew, *Sorex araneus*. *Journal of*
729   *Evolutionary Biology*, *21*(2), 636–641. doi:10.1111/j.1420-9101.2007.01481.x
730 Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish!
731   Why not publish your data too? *Intelligence, 40*(2), 73–76.
732   doi:10.1016/j.intell.2012.01.004
733 Willmore, K. E., Klingenberg, C. P., & Hallgrimsson, B. (2005). The relationship
734   between fluctuating asymmetry and environmental variance in rhesus
735   macaque skulls. *Evolution., 59*(4), 898–909.
736 Withey, D. J., & Koles, Z. J. (2008). A review of medical image segmentation:
737   Methods and available software. *International Journal of Bioelectromagnetism*,
738   *10*(3), 125–148. Retrieved from http://www.ijbem.org
739 Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating
740   repeatability. *Methods in Ecology and Evolution, 3*(1), 129–137.
741   doi:10.1111/j.2041-210X.2011.00125.x
742 Yaniv, Z., Lowekamp, B. C., Johnson, H. J., & Beare, R. (2018). SimpleITK
743   image-analysis notebooks: a collaborative environment for education and
744   reproducible research. *Journal of Digital Imaging, 31*(3), 290–303.
745   doi:10.1007/s10278-017-0037-8
746 Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics*
747   *for biologists: a primer*. New York: Academic Press.
748

749

750 **Figure legends**

751 Figure 1. Schematic diagram of the three segmentation methods that were

752   compared. (a) Distribution of the Hounsfield unit (HU) values in two bone

753   regions with high and low values, respectively. (b) Segmentation based on local

754   half-maximum height (HMH) (reference model). (c) Segmentation based on the

755   visual-based method. That the high-HU region is erroneously inflated

756   compared with the reference model, when the low-HU region is correctly

757   segmented. (d) Segmentation based on the HMH-based method. If a global

758   threshold is calculated to be higher than the local HMH of low-HU region, the

759   low-HU region could not be segmented or is erroneously eroded. (e)

760   Segmentation based on the image gradient with watershed (GWS)-based

761   method. With this approach, watersheds are searched for based on the

762   gradient magnitude by placing seeds within the catchment based on different

763   thresholds. Although some regions lacked seeds, this is often less of an issue

764 because there are usually seeds within the same catchment. The boundary of
765 segmentation is often close to local HMH in both the high- and low-HU regions.

766 Figure 2. (a–c) Three-dimensional surface reconstructions obtained using the
767 visual-based method (grey) (a), the half-maximum height (HMH)-based method
768 (orange) (b) and the image gradient with watershed (GWS)-based method (sky
769 blue) (c). (d, e) Superimposition of the visual- and GWS-based models (d) and
770 the GWS- and HMH-based models (e). (f) Pictures of the original cranium
771 showing the left-lateral (upper) and occlusal (bottom) views. The specimen is
772 PRICT-1257 (PRISK-5866).

773 Figure 3. The segmentation of the visual-based method (grey), the half-maximum
774 height (HMH)-based method (orange) and the image gradient with watershed
775 (GWS)-based methods (sky blue) in comparison to a reference model (a local
776 HMH). Positive values indicate that the model was too thick, while negative
777 values indicate too thin, compared to the reference (local HMH). Zero indicates
778 that it is identical to the reference model.

779 Figure 4. Inter-method error: g_v, between the image gradient with watershed
780 （GWS）-based method and the visual-based method; g_h, between the GWS-
781 and the half-maximum height (HMH)-based methods; h_v, between the HMH-
782 and the visual-based methods. Herein, the mean values of replicates are used.
783 Horizontal dashed lines denote the median of intra-observer errors. Outliers
784 are not shown. For bilateral landmarks, the left- and right-side data are pooled.
785 Asterisk indicates whether inter-method error is significantly larger/smaller
786 than intra-observer error (*, $P = 0.05$; **, $P = 0.01$).

787 Figure 5. The scatterplots of PC scores: (a) PC1 vs. PC2, (b) PC3 vs. PC4. Colour
788 denotes segmentation methods; symbol denotes species. The data of same
789 individual are circled by 95% confidence ellipse.

790

791 **Tables**

Table 1. Samples used in this study.

PRICT No[a]	PRISK ID[b]	Species
322	528	*M. cyclopis*
323	1358	*M. cyclopis*
342	3046	*M. fascicularis*
357	4477	*M. fascicularis*
989	8644	*M. fuscata*
1001	8658	*M. fuscata*
1005	9332	*M. fuscata*
1007	9340	*M. fuscata*
1010	9361	*M. fuscata*
1239	6162	*M. fuscata*
1243	6470	*M. fuscata*
1245	6474	*M. fuscata*
1257	5866	*M. fuscata*
325	218	*M. mulatta*
326	223	*M. mulatta*
900	1849	*M. nemestrina*
903	2299	*M. nemestrina*
387	3052	*M. radiata*
920	9532	*M. radiata*

[a] Digital Morphology Museum, KUPRI

(dmm3.pri.kyoto-u.ac.jp/dmm/WebGallery/)

[b] Skeletal collection, KUPRI

(pri.kyoto-u.ac.jp/databases/matedb/)

792

793

25

Table 2. The repeatability of HMH calculation.

Number of lines	ICC	Lower CI	Upper CI
10	0.075	-0.001	0.234
100	0.692	0.541	0.837
1,000	0.946	0.906	0.975
10,000	0.995	0.991	0.998

ICC: intraclass correlation coefficient.

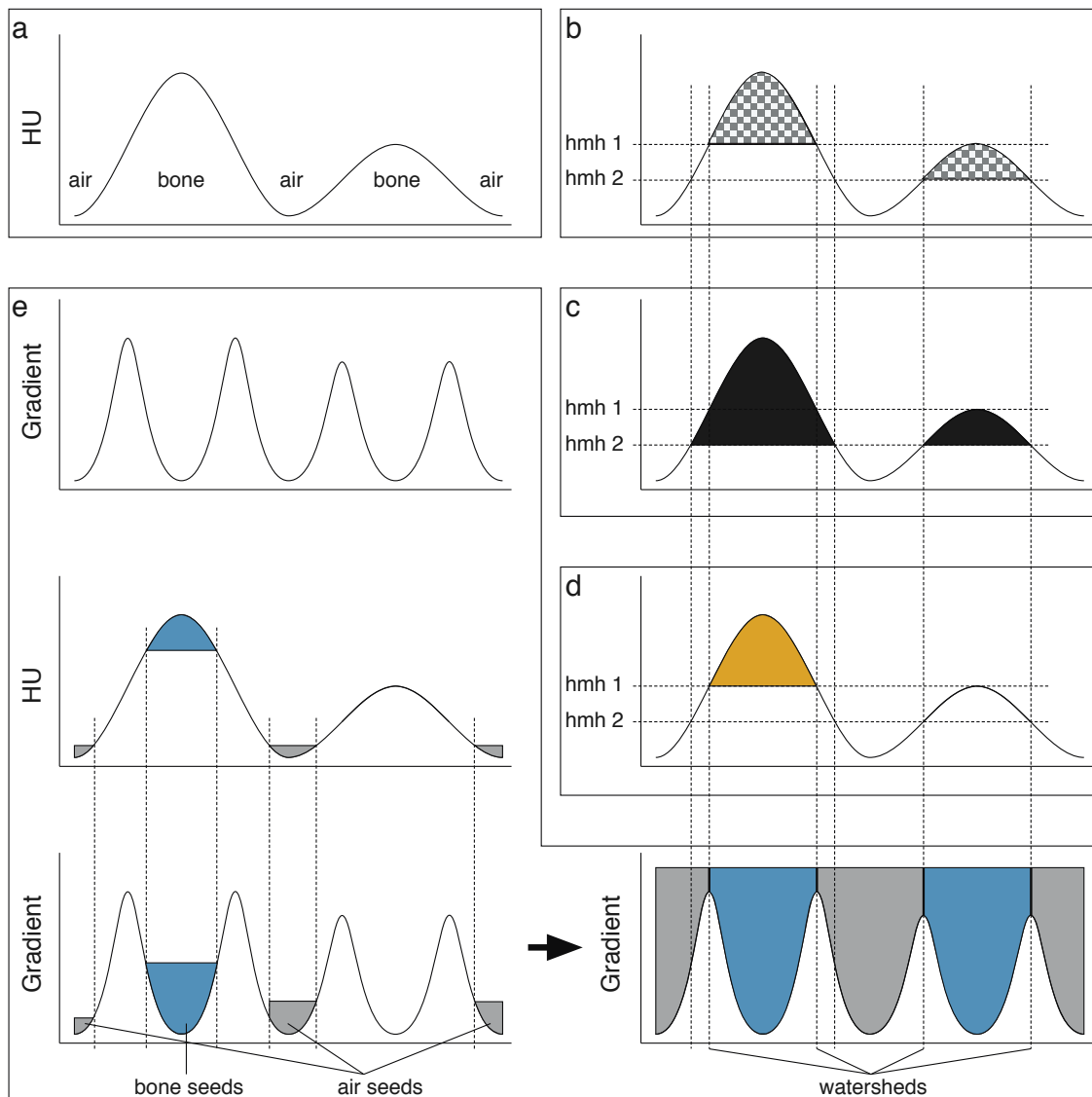CI (confidence interval limit) was estimated with the alpha level of 0.05.

794
795

Table 3. Procrustes ANOVA for centroid size and shape.

Effect	SS	MS	df	F	P
Centroid size					
Extra 1 (species)	50445	10088.98	5	21.7	<.0001
Individual	6048	465.23	13	531.3	<.0001
Error 1 (method)	33	0.88	38	9.9	<.0001
Residual	5	0.09	57		
Shape					
Extra 1 (species)	0.34898692	0.0011442	305	4.9	<.0001
Individual	0.18569714	0.0002342	793	17.9	<.0001
Side (DA)	0.00501611	0.0000965	52	7.4	<.0001
Ind * Side (FA)	0.01222904	0.0000131	936	4.1	<.0001
Error 1 (method)	0.01375109	0.0000032	4294	2.1	<.0001
Residual	0.00993181	0.0000015	6441		

DA: directional asymmetry, FA: fluctuating asymmetry.

796
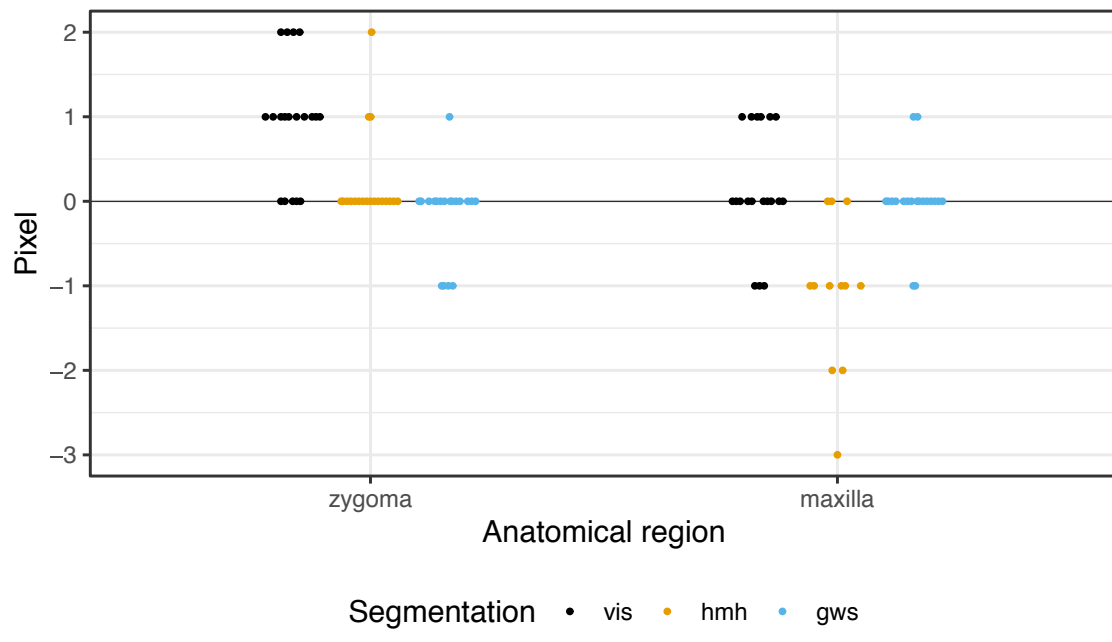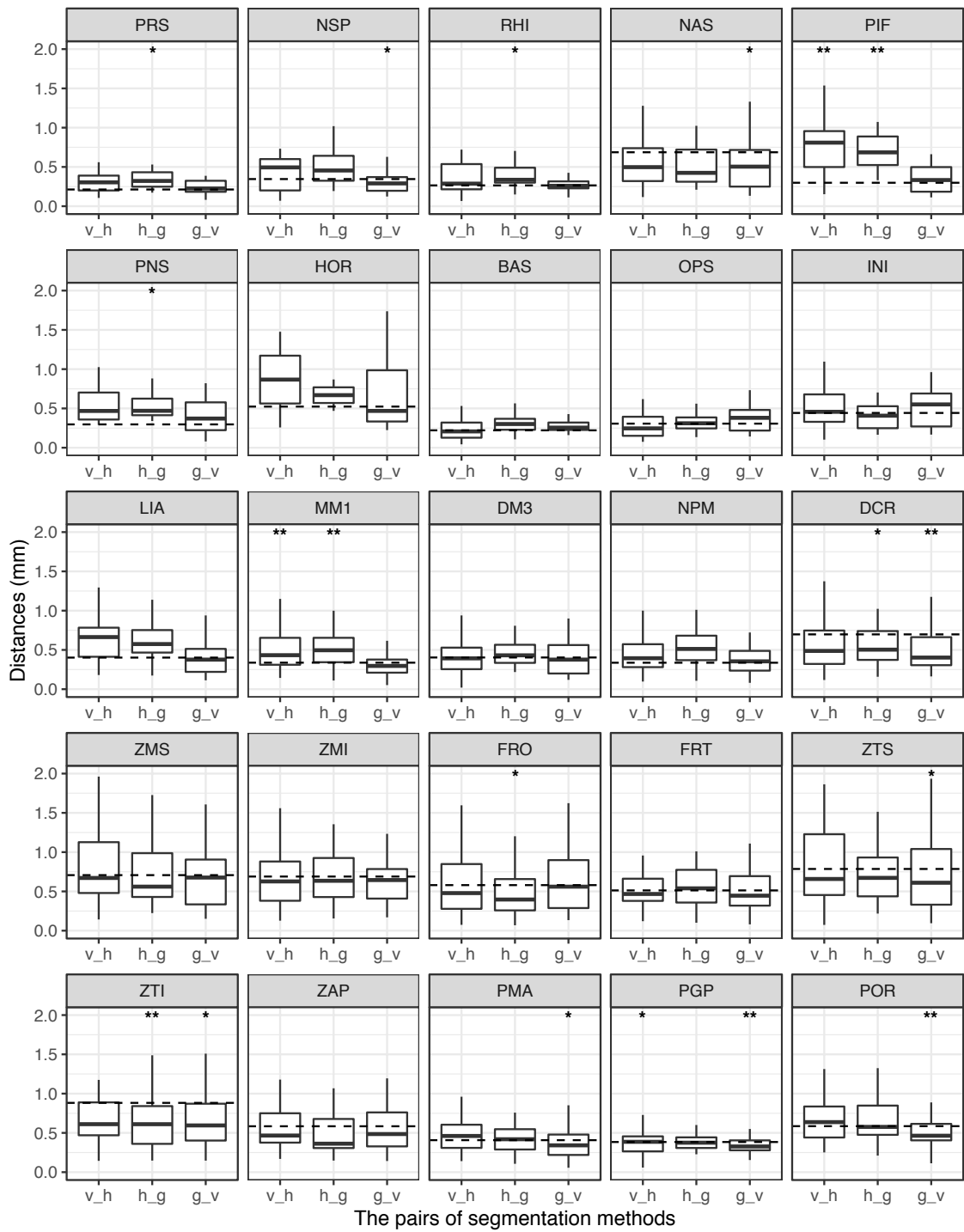797

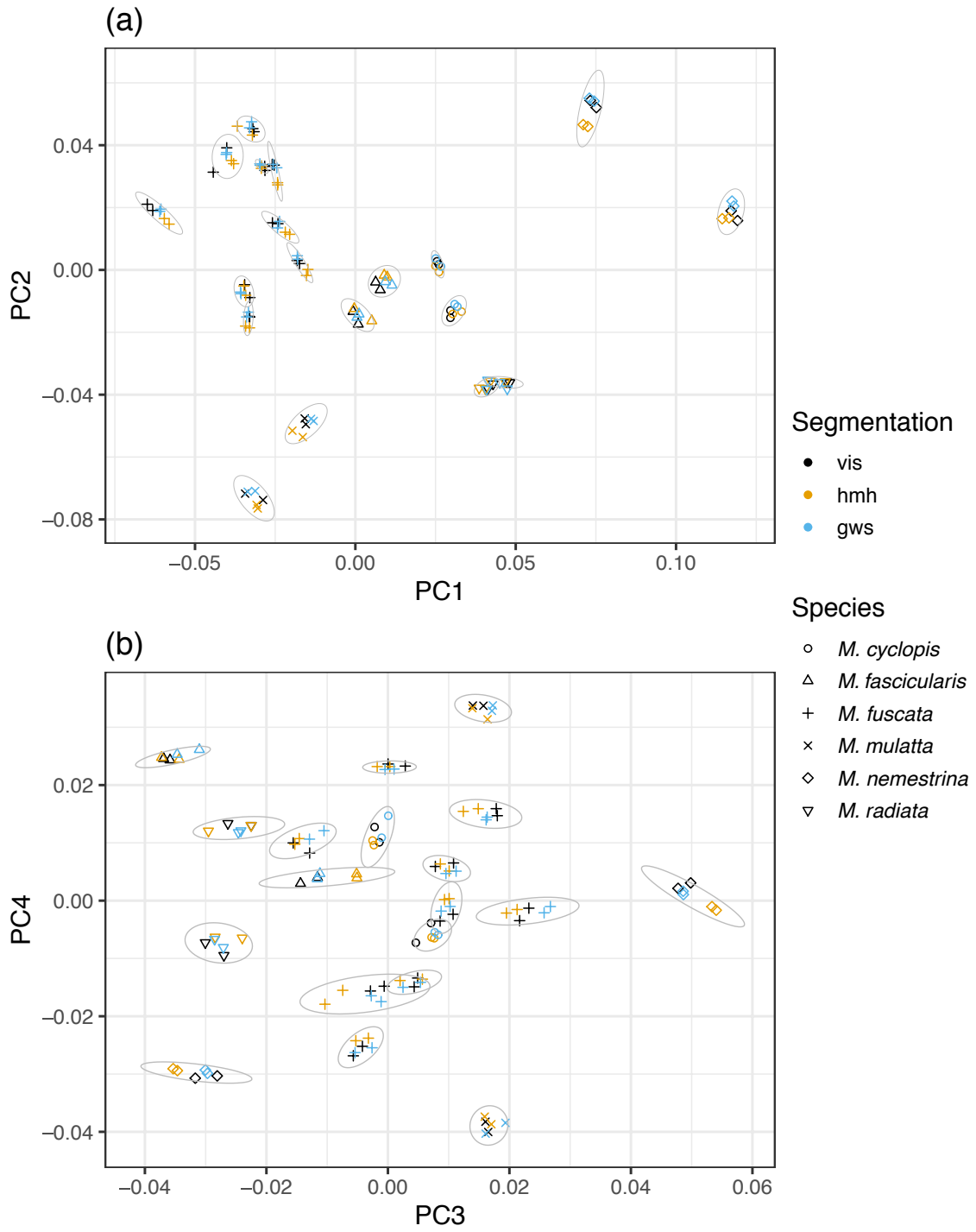**Figures**

801   Fig. 2

802

28

803   Fig. 3

804

Fig. 4

30
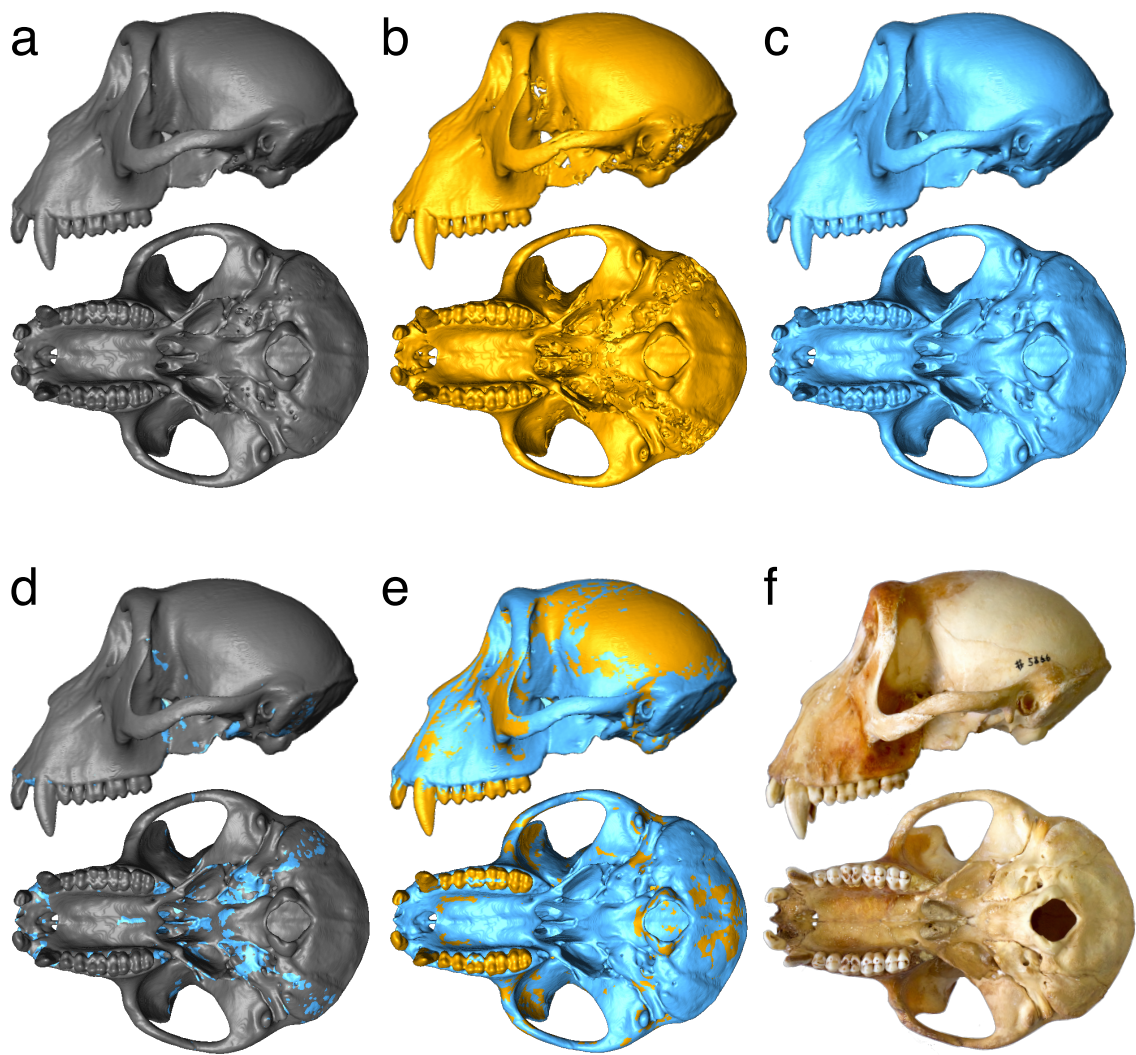
Table S1. Ecological and evolutionary insights from geometric morphometric data.	
Typical questions	Example references
Integration and modularity, response (direction) to selection	(Mitteroecker & Bookstein, 2008; Klingenberg, Debat, & Roff, 2010; Martínez-Abadías et al., 2012; Klingenberg & Marugán-Lobón, 2013)
developmental (in)stability, canalization	(Klingenberg & McIntyre, 1998; Debat, Alibert, David, Paradis, & Auffray, 2000; Santos, Iriarte, & Céspedes, 2005)
Ontogenetic and allometric trajectories and their relation to evolution	(de León & Zollikofer, 2001; Mitteroecker, Gunz, & Bookstein, 2005; Mitteroecker & Bookstein, 2009)
Evolutionary process (e.g., genetic drift vs. natural selection), tempo and mode in evolution	(Perez & Monteiro, 2009; Chira et al., 2018; Ponce de León et al., 2018; Weaver & Gunz, 2018)
Ecological, geographical, chronological and phylogenetic diversity	(Adams & Rohlf, 2000; Claude, Pritchard, Tong, Paradis, & Auffray, 2004; Cardini, Jansson, & Elton, 2007; Drake & Klingenberg, 2010; Esquerré & Keogh, 2016)
Phylogenetic inference, phylogenetic signal, taxonomic classification	(Cardini & Elton, 2008; González-José, Escapa, Neves, Cúneo, & Pucciarelli, 2008; Catalano, Goloboff, & Giannini, 2010; Catalano, Ercoli, & Prevosti, 2014; Klingenberg & Gidaszewski, 2010; Goloboff & Catalano, 2011; Détroit et al., 2019)
Genotype–phenotype mapping	(Liu et al., 2012; Pallares, Harr, Turner, & Tautz, 2014; Mitteroecker, Cheverud, & Pavlicev, 2016; Gunz et al., 2019)

References

Adams, D. C., & Rohlf, F. J. (2000). Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4106–11. doi:10.1073/pnas.97.8.4106

Cardini, A., & Elton, S. (2008). Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. *Biological Journal of the Linnean Society*, *93*(4), 813–834. doi:10.1111/j.1095-8312.2008.01011.x

Cardini, A., Jansson, A.-U., & Elton, S. (2007). A geometric morphometric approach to the study of ecogeographical and clinical variation in vervet monkeys. *Journal of Biogeography*, *34*(10), 1663–1678. doi:10.1111/j.1365-2699.2007.01731.x

Catalano, S. A., Ercoli, M. D., & Prevosti, F. J. (2014). The more, the better: The use of multiple landmark configurations to solve the phylogenetic relationships in musteloids. *Systematic Biology*, *64*(2), 294–306. doi:10.1093/sysbio/syu107

Catalano, S. A., Goloboff, P. A., & Giannini, N. P. (2010). Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework. *Cladistics.*, *26*(5), 539–549. doi:10.1111/j.1096-0031.2010.00302.x

Chira, A. M., Cooney, C. R., Bright, J. A., Capp, E. J. R., Hughes, E. C., Moody, C. J. A., … Thomas, G. H. (2018). Correlates of rate heterogeneity in avian ecomorphological traits. *Ecology Letters*, *21*(10), 1505–1514. doi:10.1111/ele.13131

Claude, J., Pritchard, P. C. H., Tong, H., Paradis, E., & Auffray, J. C. (2004). Ecological correlates and evolutionary divergence in the skull of turtles: A geometric morphometric assessment. *Systematic Biology*, *53*(6), 933–948. doi:10.1080/10635150490889498

de León, M. S. P., & Zollikofer, C. P. E. (2001). Neanderthal cranial ontogeny and its implications for late hominid diversity. *Nature*, *412*(6846), 534–538. doi:10.1038/35087573

Debat, V., Alibert, P., David, P., Paradis, E., & Auffray, J. C. (2000). Independence between developmental stability and canalization in the skull of the house mouse. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1442), 423–430. doi:10.1098/rspb.2000.1017

Détroit, F., Mijares, A. S., Corny, J., Daver, G., Zanolli, C., Dizon, E., … Piper, P. J. (2019). A new species of *Homo* from the Late Pleistocene of the Philippines. *Nature*, *568*(7751), 181–186. doi:10.1038/s41586-019-1067-9

Drake, A. G., & Klingenberg, C. P. (2010). Large- scale diversification of skull shape in domestic dogs: Disparity and modularity. *The American Naturalist*, *175*(3), 289–301. doi:10.1086/650372

Esquerré, D., & Keogh, S. J. (2016). Parallel selective pressures drive convergent diversification of phenotypes in pythons and boas. *Ecology Letters*, *19*(7), 800–809. doi:10.1111/ele.12620

Goloboff, P. A., & Catalano, S. A. (2011). Phylogenetic morphometrics (II): Algorithms for landmark optimization. *Cladistics*, *27*(1), 42–51. doi:10.1111/j.1096-0031.2010.00318.x

González-José, R., Escapa, I., Neves, W. A., Cúneo, R., & Pucciarelli, H. M. (2008). Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. *Nature*, *453*(7196), 775–778. doi:10.1038/nature06891

Gunz, P., Tilot, A. K., Wittfeld, K., Teumer, A., Shapland, C. Y., van Erp, T. G. M., … Fisher, S. E. (2019). Neandertal introgression sheds light on modern human endocranial globularity. *Current Biology*, *29*(1), 120-127.e5.

doi:10.1016/j.cub.2018.10.065

Klingenberg, C. P., Debat, V., & Roff, D. A. (2010). Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. *Evolution*, *29*, 2935–2951. doi:10.1111/j.1558-5646.2010.01030.x

Klingenberg, C. P., & Gidaszewski, N. A. (2010). Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Systematic Biology*, *59*(3), 245–261. doi:10.1093/sysbio/syp106

Klingenberg, C. P., & Marugán-Lobón, J. (2013). Evolutionary covariation in geometric morphometric data: Analyzing integration, modularity, and allometry in a phylogenetic context. *Systematic Biology*, *62*(4), 591–610. doi:10.1093/sysbio/syt025

Klingenberg, C. P., & McIntyre, G. S. (1998). Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*, *52*(5), 1363–1375.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., … Kayser, M. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genetics*, *8*(9), e1002932. doi:10.1371/journal.pgen.1002932

Martínez-Abadías, N., Esparza, M., Sjøvold, T., González-José, R., Santos, M., Hernández, M., … Klingenberg, C. P. (2012). Pervasive genetic integration directs the evolution of human skull shape. *Evolution*, *66*(4), 1010–1023. doi:10.1111/j.1558-5646.2011.01496.x

Mitteroecker, P., & Bookstein, F. (2008). The evolutionary role of modularity and integration in the hominoid cranium. *Evolution*, *62*(4), 943–958. doi:10.1111/j.1558-5646.2008.00321.x

Mitteroecker, P., & Bookstein, F. (2009). The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution*, *63*(3), 727–737. doi:10.1111/j.1558-5646.2008.00587.x

Mitteroecker, P., Cheverud, J. M., & Pavlicev, M. (2016). Multivariate analysis of genotype-phenotype association. *Genetics*, *202*(4), 1345–1363. doi:10.1534/genetics.115.181339

Mitteroecker, P., Gunz, P., & Bookstein, F. L. (2005). Heterochrony and geometric morphometrics: a comparison of cranial growth in *Pan paniscus* versus *Pan troglodytes*. *Evolution & Development*, *7*, 244–258. doi:10.1111/j.1525-142X.2005.05027.x

Pallares, L. F., Harr, B., Turner, L. M., & Tautz, D. (2014). Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. *Molecular Ecology*, *23*(23), 5756–5770. doi:10.1111/mec.12968

Perez, S. I., & Monteiro, L. R. (2009). Nonrandom factors in modern human morphological diversification: A study of craniofacial variation in southern South American populations. *Evolution*, *63*(4), 978–993. doi:10.1111/j.1558-5646.2008.00539.x

Ponce de León, M. S., Koesbardiati, T., Weissmann, J. D., Milella, M., Reyna-Blanco, C. S., Suwa, G., … Zollikofer, C. P. E. (2018). Human bony labyrinth is an indicator of population history and dispersal from Africa. *Proceedings of the National Academy of Sciences*, *115*(23), 201717873. doi:10.1073/pnas.1717873115

Santos, M., Iriarte, P. F., & Céspedes, W. (2005). Genetics and geometry of canalization and developmental stability in *Drosophila subobscura*. *BMC Evolutionary Biology*, *5*(1), 7. doi:10.1186/1471-2148-5-7

Weaver, T. D., & Gunz, P. (2018). Using geometric morphometric visualizations of directional selection gradients to investigate morphological differentiation. *Evolution*, *72*(4), 838–850. doi:10.1111/evo.13460

Table S2. Landmarks used in this study.

No		Abbreviatio	Definition	Reference
1		PRS	Prosthion: antero-inferior point on projection of premaxilla between central	Cardini et al. (2007)
2		NSP	Nasospinale: inferior-most midline point of piriform aperture.	Cardini et al. (2007)
3		RHI	Rhinion: most anterior midline point on nasals.	Cardini et al. (2007)
4		NAS	Nasion: midline point on fronto-nasal suture.	Cardini et al. (2007)
5		PIF	Posterior-most point of incisive foramen.	Cardini et al. (2007)
6		PNS	Tip of posterior nasal spine.	Cardini et al. (2007)
7		HOR	Hormion: midpoint of the posterosuperior border of the vomer.	Cobb and O'Higgins (2007)
8		BAS	Basion: anterior most point of foramen magnum.	Frost et al. (2003)
9		OPS	Opisthion: posterior most point of foramen magnum.	Frost et al. (2003)
10		INI	Inion: most posterior point of cranium, when viewed in the Frankfurt horizontal, be it on sagittal/nuchal crest or not.	Frost et al. (2003)
11	26	LIA	Posterior-most point of lateral incisor alveolus.	Cardini et al. (2007)
12	27	MM1	Mesial M1: contact points between P^4 and M^1, projected labially onto alveolar	Cardini et al. (2007)
13	28	DM3	Distal M3: posterior midpoint onto alveolar margin of M^3.	Cardini et al. (2007)
14	29	NPM	Meeting point of nasal and pre-maxilla on margin of piriform aperture.	Cardini et al. (2007)
15	30	DCR	Dacryon: most superior point of the lacrimomaxillary suture (intersection with frontal bone).	Cobb and O'Higgins (2007)
16	31	ZMS	Zygo-max superior: antero-superior point of zygomaticomaxillary suture taken at orbit rim.	Cardini et al. (2007)
17	32	ZMI	Zygo-max inferior: antero-inferior point of zygomaticomaxillary suture.	Cardini et al. (2007)
18	33	FRO	Frontomalare orbitale: where frontozygomatic suture crosses inner orbital rim.	Cardini et al. (2007)
19	34	FRT	Frontomalare temporale: where frontozygomatic suture crosses lateral edge of	Cardini et al. (2007)
20	35	ZTS	Zygo-temp superior: superior point of zygomatico-temporal suture on lateral face of zygomatic arch.	Frost et al. (2003)
21	36	ZTI	Zygo-temp inferior: inferior point of zygomatico-temporal suture on lateral face of zygomatic arch.	Frost et al. (2003)
22	37	ZAP	Meeting point of zygomatic arch and alisphenoid on superior margin of pterygomaxillary fissure.	Cardini et al. (2007)
23	38	PMA	Most posterior point of maxillary alveolus on the maxilla-palatine.	Ito et al. (2014)
24	39	PGP	Most inferior point on the postglenoid process.	Lockwood et al. (2004)
25	40	POR	Porion: top of auditory meatus.	Frost et al. (2003)

Table S3. Inter-method and intra-observer errors for each landmark (mm) .

	visual–HMH					HMH–GWS					GWS–visual					Intra-observer		
	Mean	Median	SD	*P*	*	Mean	Median	SD	*P*	*	Mean	Median	SD	*P*	*	Mean	Median	SD
PRS	0.32	0.30	0.18	0.15	+	0.36	0.32	0.14	**0.01**	+	0.25	0.23	0.09	0.78	-	0.26	0.21	0.15
NSP	0.47	0.49	0.28	0.80	+	0.55	0.45	0.36	0.28	+	0.32	0.29	0.18	**0.05**	-	0.45	0.35	0.34
RHI	0.37	0.29	0.21	0.11	+	0.41	0.34	0.18	**0.02**	+	0.28	0.26	0.12	0.80	+	0.27	0.27	0.15
NAS	0.65	0.51	0.58	0.24	-	0.63	0.44	0.50	0.16	-	0.55	0.50	0.36	**0.02**	-	0.83	0.69	0.60
PIF	0.75	0.81	0.35	**6.E-04**	+	0.68	0.69	0.25	**1.E-04**	+	0.38	0.33	0.25	0.48	+	0.33	0.30	0.23
PNS	0.58	0.47	0.32	0.08	+	0.55	0.47	0.20	**0.04**	+	0.40	0.37	0.23	0.97	-	0.40	0.30	0.31
HOR	0.87	0.87	0.86	0.74	+	0.67	0.67	0.28	0.80	+	0.96	0.59	0.89	0.12	+	0.60	0.52	0.41
BAS	0.23	0.21	0.14	0.37	-	0.31	0.30	0.12	0.31	+	0.28	0.26	0.07	0.84	+	0.27	0.22	0.17
OPS	0.28	0.25	0.16	0.17	-	0.33	0.31	0.13	0.79	-	0.37	0.38	0.17	0.54	+	0.34	0.31	0.22
INI	0.50	0.46	0.25	0.51	-	0.48	0.41	0.33	0.42	-	0.51	0.55	0.26	0.59	-	0.55	0.44	0.41
LIA	0.70	0.67	0.44	0.45	+	0.85	0.59	1.12	0.25	+	0.59	0.38	1.16	1.00	+	0.59	0.40	1.37
MM1	0.50	0.43	0.26	**3.E-03**	+	0.52	0.50	0.25	**5.E-04**	+	0.31	0.30	0.13	0.12	-	0.36	0.34	0.17
DM3	0.42	0.39	0.27	0.75	-	0.49	0.43	0.24	0.28	+	0.41	0.38	0.22	0.54	-	0.44	0.40	0.26
NPM	0.46	0.39	0.28	0.71	+	0.52	0.51	0.23	0.06	+	0.41	0.35	0.26	0.61	-	0.43	0.34	0.32
DCR	0.59	0.49	0.37	0.06	-	0.57	0.50	0.30	**0.01**	-	0.52	0.40	0.32	**2.E-03**	-	0.72	0.70	0.35
ZMS	0.83	0.67	0.46	0.96	-	0.76	0.56	0.46	0.47	-	0.76	0.69	0.50	0.44	-	0.83	0.71	0.54
ZMI	0.67	0.63	0.35	0.17	-	0.71	0.64	0.31	0.37	-	0.66	0.64	0.32	0.15	-	0.76	0.69	0.46
FRO	0.57	0.48	0.38	0.26	-	0.50	0.40	0.35	**0.03**	-	0.66	0.56	0.42	0.86	+	0.65	0.58	0.43
FRT	0.60	0.48	0.43	0.71	-	0.60	0.54	0.34	0.62	-	0.52	0.45	0.26	0.06	-	0.63	0.51	0.49
ZTS	0.94	0.75	0.62	0.82	-	0.77	0.68	0.45	0.06	-	0.71	0.61	0.49	**0.02**	-	0.96	0.79	0.69
ZTI	0.87	0.62	0.62	0.18	-	0.67	0.61	0.44	**1.E-03**	-	0.77	0.62	0.51	**0.02**	-	1.06	0.88	0.89
ZAP	0.56	0.47	0.26	0.21	-	0.52	0.36	0.33	0.10	-	0.55	0.49	0.31	0.20	-	0.63	0.58	0.38
PMA	0.47	0.46	0.20	0.66	+	0.43	0.42	0.20	0.68	-	0.37	0.34	0.18	**0.04**	-	0.45	0.41	0.29
PGP	0.38	0.39	0.16	**0.05**	-	0.39	0.38	0.14	0.08	-	0.35	0.33	0.12	**1.E-03**	-	0.46	0.38	0.27
POR	0.68	0.64	0.33	0.81	+	0.67	0.58	0.30	0.99	-	0.52	0.46	0.23	**0.01**	-	0.67	0.59	0.43

* Plus/minus indicates that the mean of inter-method error is larger/smaller than the mean of intra-observer error.
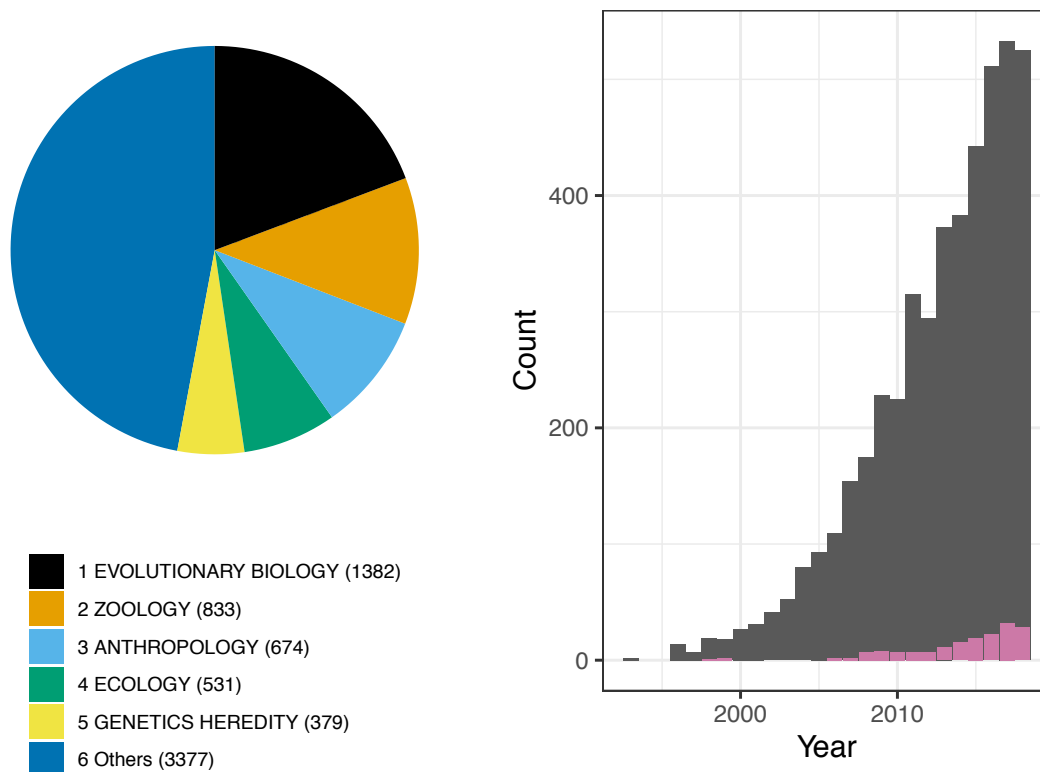
Figure S1. Literature published between 1990 and 2018 that is related to geometric morphometrics. Publications were retrieved from Web of Science using the search term 'geometric morphometric$'. The proportion of papers in each research field is shown on the left based on the tags provided in Web of Science, while the number of publications per year is shown on the right. The purple bar denotes the number of publications retrieved when 'computed tomography' was added to the search term. Note that the Web of Science topic search only scans the title, abstract, author keywords and keywords plus. By contrast, Google Scholar detected approximately 12,900 publications over the same period (or 1,640 when 'computed tomography' was added).
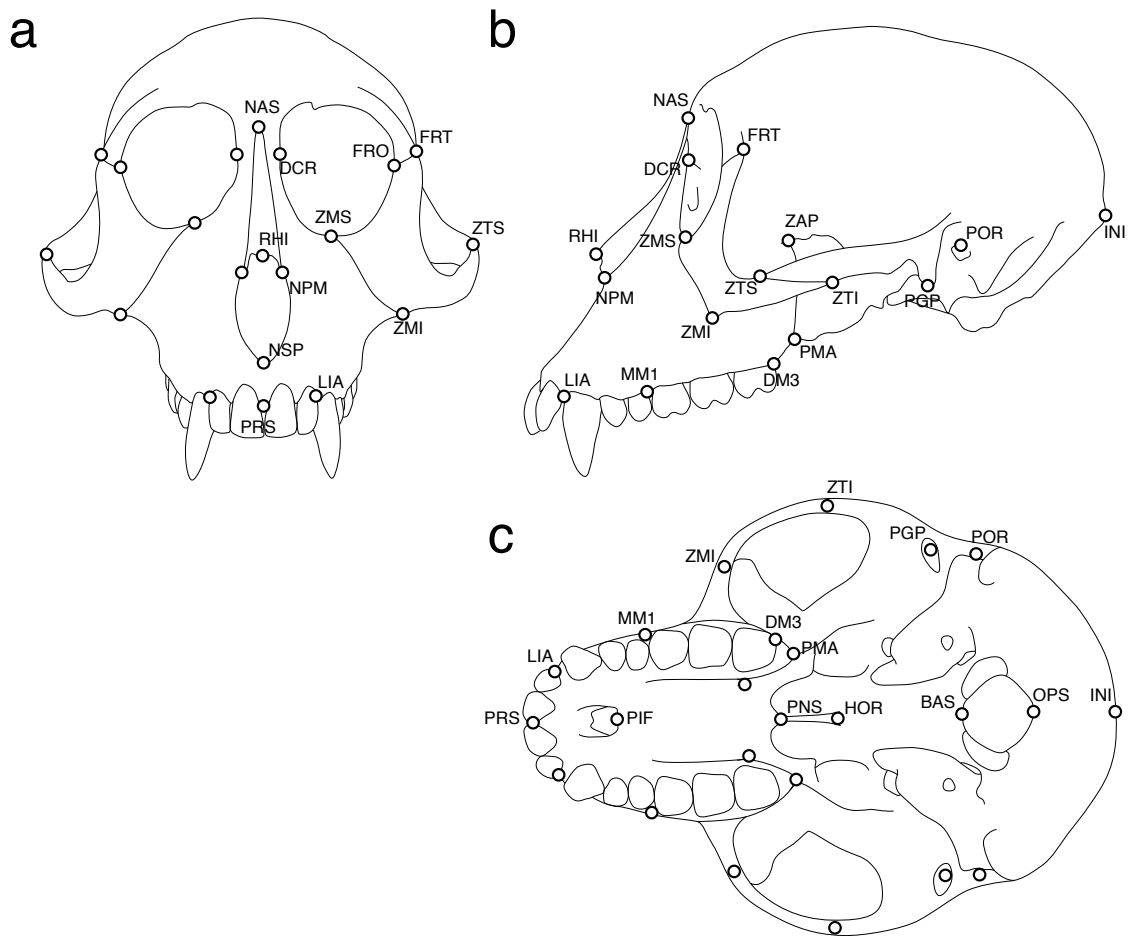
Figure S2. Landmarks used in this study. (a) frontal, (b) left-lateral, and (c) occlusal views of cranium. See also Table S2.
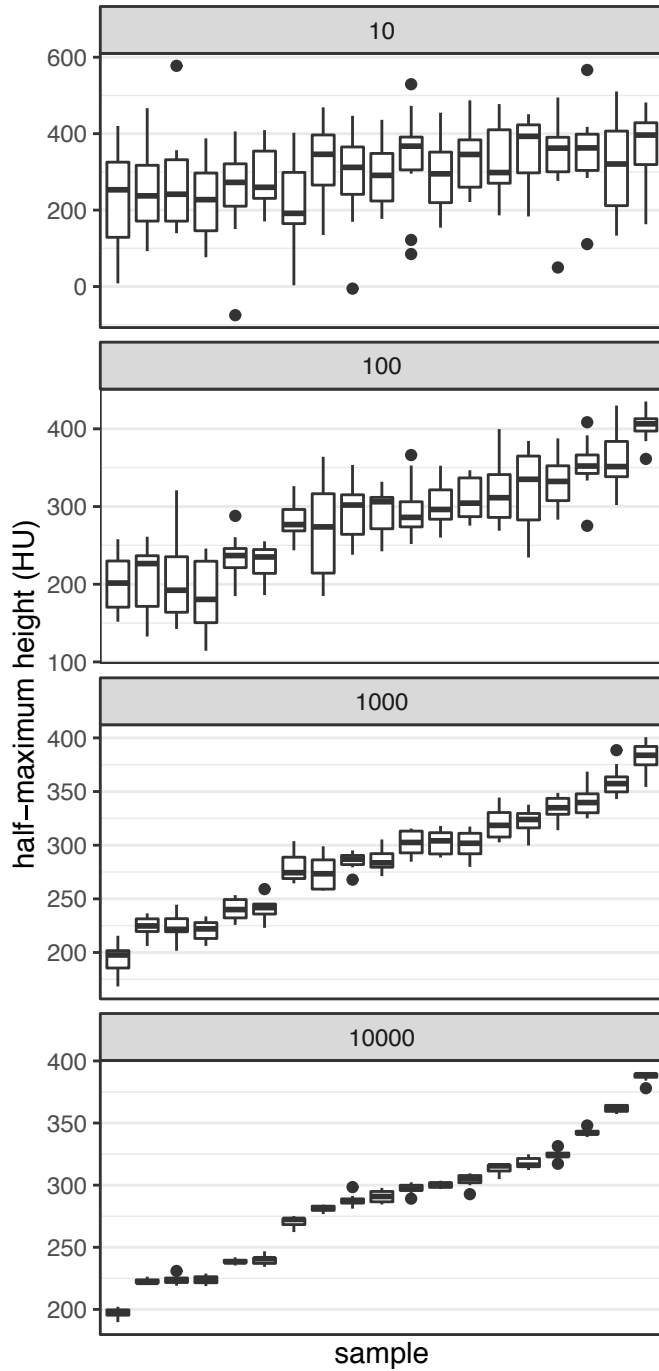
Figure S3. The repeatability of the half-maximum height (HMH)-based methods relying on the 10, 100, 1,000, and 10,000 randomly drawn lines. Horizontal axis indicates samples: from the left, PRICT-325, 1257, 357, 322, 900, 1243, 920, 342, 989, 1001, 326, 1010, 1007, 387, 1005, 1245, 903, 323 and 1239.
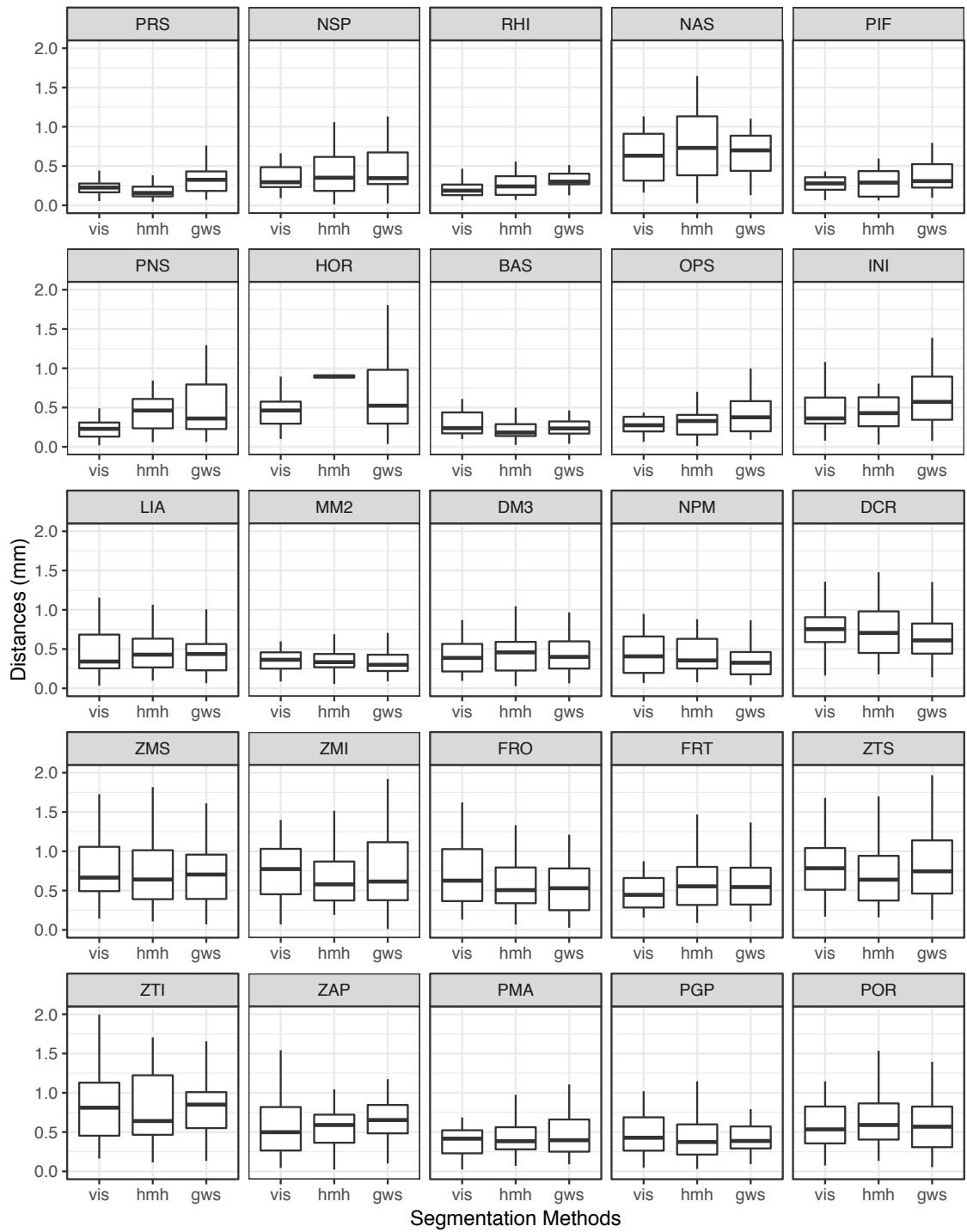
Figure S4. Intra-observer error: vis, the visual-based model; hmh, the half-maximum height (HMH)-based method; gws, the image gradient with watershed (GWS)-based model. For bilateral landmarks, the left- and right-side data are pooled. Outliers are not shown.