High-density morphometric analysis of shape and integration: the good, the bad, and the not-really-a-problem

SCHOLARONE™
Manuscripts

1 High-density morphometric analysis of shape and integration: the good, the bad, and

2 the not-really-a-problem

3 Running title: Big data analysis of shape and integration

4

5 Anjali Goswami[1,2*], Akinobu Watanabe[1,3,4], Ryan N. Felice[1,5], Carla Bardua[1,2], Anne-

6 Claire Fabre[1], P. David Polly[6]

7 　[1]Life Sciences Department, Vertebrates Division, Natural History Museum, London, SW7

8 5BD, UK

9 　[2]Department of Genetics, Evolution and Environment, University College London, London

10 WC1E 6BT, UK

11 　[3]Department of Anatomy, New York Institute of Technology College of Osteopathic

12 Medicine, Old Westbury, NY 11568, USA

13 　[4]Division of Paleontology, American Museum of Natural History, New York, NY 10024, USA

14 　[5]Centre for Integrative Anatomy, Department of Cell and Developmental Biology, University

15 College London, London WC1E 6BT, UK

16 　[6]Departments of Geological Sciences, Biology, and Anthropology, Indiana University, 1001

17 E. 10th Street, Bloomington, IN 47405, USA.

18 　*Corresponding author, a.goswami@nhm.ac.uk; +44 (0)20 7679 5063

19

Abstract

20

21 The field of comparative morphology has entered a new phase with the rapid generation

22 of high-resolution three-dimensional data. With freely available 3D data of thousands of

23 species, methods for quantifying morphology that harness this rich phenotypic

24 information are quickly emerging. Among these techniques, high-density geometric

25 morphometric approaches provide a powerful and versatile framework to robustly

26 characterize shape and phenotypic integration, the covariances among morphological

27 traits. These methods are particularly useful for analyses of complex structures and

28 across disparate taxa, which may share few landmarks of unambiguous homology.

29 However, high-density geometric morphometrics also brings challenges, for example

30 with statistical, but not biological, covariances imposed by placement and sliding of

31 semilandmarks and registration methods such as Procrustes superimposition. Here, we

32 present simulations and case studies of high-density datasets for squamates, birds, and

33 caecilians that exemplify the promise and challenges of high-dimensional analyses of

34 phenotypic integration and modularity. We assess: (1) the relative merits of "big" high-

35 density geometric morphometrics data over traditional shape data; (2) the impact of

36 Procrustes superimposition on analyses of integration and modularity; and (3)

37 differences in patterns of integration between analyses using high-density geometric

38 morphometrics and those using discrete landmarks. We demonstrate that for many skull

39 regions 20-30 landmarks and/or semilandmarks are needed to accurately characterize

40 their shape variation, and landmark-only analyses do a particularly poor job of capturing

41 shape variation in vault and rostrum bones. Procrustes superimposition can mask

42 modularity, especially when the number of landmarks is low and they covary in parallel

43   directions, but this effect decreases with increasing landmark number or more

44   biologically complex covariance patterns. Landmark-only and landmark-plus-sliding-

45   semilandmark analyses of integration are generally congruent in overall pattern of

46   integration, but landmark-only analyses tend to show higher integration between

47   adjacent bones, especially when landmarks placed on the sutures between bones

48   introduces a boundary bias. Allometry may be a stronger influence on patterns of

49   integration in landmark-only analyses, which show stronger integration prior to removal

50   of allometric effects compared to analyses including semilandmarks. High-density

51   geometric morphometrics has its challenges and drawbacks, but our analyses of

52   simulated and empirical datasets demonstrate that these potential issues are unlikely to

53   obscure genuine biological signal. Rather, high-density geometric morphometric data

54   exceeds traditional landmark-based methods in characterization of morphology and

55   allow more nuanced comparisons across disparate taxa. Combined with the rapid

56   increases in 3D data availability, high-density morphometric approaches have immense

57   potential to propel a new class of studies of comparative morphology and phenotypic

58   integration.

59

Introduction

60

61   Big data approaches to morphological studies have entered a new phase in recent

62   years, due to the ubiquity of high-resolution imaging tools, such as micro-CT imaging

63   and surface scanning and photogrammetry (Davies *et al.* 2017). Open databases

64   (Morphosource, Phenome10K, Digimorph, Morphomuseum, and institutional sites) now

65   host 3D image files for tens of thousands of specimens, meaning that obtaining access

66   to 3D scans representing a substantial proportion of the extant, and even extinct

67   diversity, for clades as large as all vertebrates, is rapidly become the expectation, rather

68   than a pipe dream. These new datasets open new possibilities for investigating

69   biological questions (Collyer *et al.* 2014), including comparative analyses that can begin

70   to quantify and analyse morphology at an extremely high level of detail across wider

71   taxonomic scales (Fig. 1).

72   To date, most comparative studies using geometric morphometrics (GM)

73   comparing morphology in a quantitative framework have either sampled closely related

74   taxa that share substantial numbers of landmarks of unambiguous homology (i.e., Type

75   I/II landmarks following Bookstein (1991)) or sample a broader taxonomic scope but

76   using a much reduced number of landmarks. Alternatively, analyses may use traditional

77   metrics, such as linear measurements, which capture some aspect of the morphology of

78   functionally analogous regions (e.g. rostrum) that can be compared directly across

79   diverse taxa, but provide very limited detail on morphology and cannot be used to

80   reconstruct shape (Marugán-Lobón and Buscalioni 2003). Recent years have seen

81   development and refinement of geometric morphometric expansions of alternatives to

82   homologous landmarks (Bookstein 1991), with application of 3D sliding semilandmarks

83   or pseudolandmarks. Published definitions of semilandmarks and pseudolandmarks are

84   inconsistent and often interchangeable, but here, we refer to semilandmarks as those

85   whose initial position is relative to landmarks with biological homology, whereas

86   pseudolandmarks are entirely automatically placed without reference to anatomically

87   defined landmarks, for example from sampling uniformly from a surface mesh (e.g.,

88   auto3dgm, Boyer *et al.* 2015; Generalized Procrustes Surface Analysis, Pomidor *et al.*

89   2016). Detailed descriptions, discussions, and comparisons of these methods (Adams

90   *et al.* 2004; Adams *et al.* 2013; Bardua *et al.* 2019a; Bookstein *et al.* 2002; Boyer *et al.*

91   2015; Gonzales *et al.* 2016; Gunz and Mitteroecker 2013; Gunz *et al.* 2005;

92   Mitteroecker and Gunz 2009; Rohlf and Marcus 1993; Vitek *et al.* 2017; Zelditch *et al.*

93   2004) demonstrate the promise these methods offer for quantifying regions that are

94   poorly characterized by use of only discrete landmarks, due to the lack of unambiguous

95   homology across specimens or the presence of large areas without any appropriate

96   structures at which to place landmarks. The lack of points of unambiguous homology

97   becomes increasingly challenging with comparative studies across large clades. For

98   example, ongoing work by our research team on tetrapod skulls identified a total of 12

99   Type I landmarks that could be reliably placed across the full cranial diversity of that

100   clade, meaning that the vast majority of cranial morphology would go unsampled (Fig.

101   1). Even for less speciose clades, such as the 32 extant genera of caecilian

102   amphibians, this can be a highly limiting factor due to a large degree of variation in bone

103   presence and suture patterns (Bardua *et al.* 2019b). The second point is an issue at

104   any scale of analysis, as many structures will only have discrete points, such as

105   sutures, at their boundaries, meaning that most of the shape of the structure will be

106   unsampled. For example, even in a clade with relatively conserved morphology such as

107   birds, a high degree of bone fusion has limited previous studies to a small number of

108   landmarks (e.g., 11-17 landmarks in Bright *et al.* 2016; Klingenberg and Marugan-Lobon

109   2013) (Fig 1).

110   While semilandmarks and pseudolandmarks are now frequently deployed to

111   circumvent these landmark-only issues (Polly 2008), questions have been raised about

112   their necessity and applicability for the study of phenotypic integration and other topics

113   in which the covariance structure of shape data is important (Cardini 2018; Lele and

114   Richtsmeier 1990; Richtsmeier and Lele 2001). Phenotypic integration refers to the

115   correlation or covariance of traits due to genetic, developmental, or functional

116   interactions (Olson and Miller 1958), and analysis of these relationships among traits

117   relies on accurate quantification of their morphology and their correlations or

118   covariances. Pseudolandmarks have not yet been used in studies of integration, and

119   their use in such studies is likely hindered by their lack of reference to biological

120   homology. In contrast, many studies have used semilandmarks to quantify the

121   relationships among different elements or regions of structures ranging from the

122   vertebrate skulls and mandibles (e.g.,Bardua *et al.* 2019a; Bardua *et al.* 2019b; Felice

123   and Goswami 2018; Marshall *et al.* 2019; Parr *et al.* 2016; Watanabe *et al.* 2019;

124   Zelditch *et al.* 2009) to fish fins (Du *et al.* 2018; Larouche *et al.* 2018) to trilobite cranidia

125   (Webster and Zelditch 2011). For this reason, we focus here on the use of

126   semilandmarks (and more specifically, sliding semilandmarks) in studies of phenotypic

127   integration, and more broadly, on their contribution to comparative studies of

128   morphological evolution.

129   The concerns about using semilandmarks for such analyses fall into two

130   categories. First, and most broadly, all geometric morphometric data, including Type I/II

131   landmarks as well as semilandmarks, require registration prior to analysis in order to

132   remove the non-shape aspects of position, orientation, and isometric size. The most

133   common method of registering specimens is generalized Procrustes superimposition

134   (Rohlf 1990; Rohlf and Slice 1990), which is a least-squares approach that mininizes

135   variance across an entire landmark (and/or semilandmark) configuration and rescales

136   each configuration to unit centroid size. Because this approach minimizes variance

137   across the entire configuration, it can have the effect of spreading variance across

138   landmarks. In other words, it may shift variance from more variable landmarks to less

139   variable ones and imposes a common scaling on a structure that may have differential

140   scaling in different regions (Baab 2013; Klingenberg 2009), both of which can alter the

141   covariance structure of the landmarks and change the inferred pattern of integration

142   among traits. It has been recently asserted that this effect may be exacerbated in larger

143   geometric morphometric datasets, such as those generated through the application of

144   semilandmarks, although such an effect was not demonstrated, and assumed that the

145   effects would reduce the ability to detect biological modularity in data (Cardini 2018).

146   Second, and more specifically, it has also been asserted that closely packed

147   semilandmarks may falsely inflate the pattern of modularity (the division of structures

148   into highly-integrated, but semi-independent subunits) because the position of each

149   semilandmark is conditional on its neighbours and therefore multiplication of

150   semilandmarks could increase the total covariance within a putative module. For these

151   reasons, it has been suggested that "big data" is not necessarily better data when it

152   comes to geometric morphometric analyses, especially analyses of phenotypic

153   integration and modularity (Cardini 2018).

154   Here, we examine these issues and their potential impact on phenomic analyses

155   of phenotypic integration. To do so, we first assess whether the gains are worth these

156 potential drawbacks by considering: 1) do high-density semilandmark datasets actually

157 capture shape better than Type I/II landmarks data?. If so, we then consider the

158 practical consequences of using these high-density data, or geometric morphometric

159 more generally, for analyses of phenotypic integration, by addressing: 2) does

160 Procrustes superimposition mislead analyses of phenotypic integration and modularity;

161 and 3) how do analyses of integration with high-dimensional semilandmarks compare to

162 those with only landmarks?

163

164 ***The effect of high-density geometric morphometric data on shape analyses***

165 To quantify whether high-density semilandmark data adds important additional

166 information on morphology, we analysed two datasets. The first dataset is from recently

167 published study of the cranium of caecilian amphibians (Fig 2A, B), with 16 crania

168 regions quantified across 32 genera using 53 landmarks and 687 curve and 729 surface

169 sliding semilandmarks (Bardua *et al.* 2019b). The second is a recently published

170 dataset of squamates (Fig 2D, E), with 13 cranial regions quantified in 174 species with

171 47 landmarks and 595 curve and 580 surface sliding semilandmarks (Watanabe *et al.*

172 2019). To examine how many landmarks/semilandmarks are required to capture the

173 shape of a region in these datasets, we implemented Landmark Sampling Evaluation

174 Curve (LaSEC) analysis, using the 'lasec' function in the R package LaMDBA

175 (Watanabe 2018). This function subsamples the original dataset through random

176 addition of landmarks and semilandmarks, determining the fit of each reduced dataset

177 to the complete dataset, and repeating this for a selected number of iterations. Fit is

178   calculated based on Procrustes distance between the full and subsampled datasets with

179   respect to position of the specimens in high-dimensional morphospace (i.e., not position

180   of the landmarks). We performed LaSEC for 1) landmarks-only and 2) subsampled

181   landmarks and semilandmarks (curve and surface points) for the caecilian and

182   squamate datasets, for individual cranial regions. The function generates a sampling

183   curve (Fig. 2C, F), where a plateau in the curve signifies stationarity in characterization

184   of shape variation and fewer landmarks than the plateau indicates inadequate

185   characterization. We compared the fit of the landmark-only and full datasets and also

186   determined the number of landmarks and semilandmarks that would have been

187   sufficient for each region, given a required fit of 0.9, 0.95 and 0.99 between the reduced

188   and complete datasets (Tables 1 and 2). To compare the relative contribution of curve

189   and surface semilandmarks to shape characterization, we further conducted LaSEC

190   analysis comparing the fit of landmarks and curve sliding semilandmarks to the full

191   dataset of landmarks and curve and surface sliding semilandmarks for the squamate

192   dataset.

193   These analyses demonstrate that landmark-only datasets do not fully capture the

194   variation of these analysed structures, with the fit between landmark-only and full

195   landmark + semilandmark datasets ranging between 0.24 to 0.81 for individual cranial

196   regions. To achieve a fit of 0.95 to a high-density dataset, cranial regions need to be

197   sampled by >20 landmarks and semilandmarks. While this cannot distinguish between

198   the value of large numbers of landmarks and similarly large numbers of curve and/or

199   surface sliding semilandmarks, it is uncontroversial that semilandmarks can sample

200   more morphology than Type I/II landmarks. In these datasets, for example, our attempt

201   to maximize representation of cranial structures with Type I/II landmarks resulted in 2-7

202   landmarks sampled per region, in comparison to the >20 landmarks and semilandmarks

203   that our analyses estimated are needed to represent the variation in each region. Thus,

204   landmark data alone are insufficient to fully characterize morphological variation for

205   many datasets. In terms of the respective contribution of curve and surface sliding

206   semilandmarks to characterizing variation, the addition of curve sliding semilandmarks

207   alone is a vast improvement on landmark-only analyses, with a fit of over 0.9 for all

208   cranial regions in squamates and approaching a near perfect fit to the full dataset for

209   relatively flat structures. However, it is important to note that the reason a similar

210   analysis would be less informative, and thus was not conducted, for the caecilian

211   dataset, is that some of the most variable regions, including the maxilla and pteryoid,

212   required the use of non-homologous curves to accommodate variably present

213   structures, such as the tentacular canal (Bardua *et al*. 2019a; Bardua *et al*. 2019b).

214   These curves were then excluded, with only landmarks and surface sliding

215   semilandmarks used in further analyses. Thus, although curves may capture much of

216   the morphological variation of the full landmark, curve, and surface dataset for many

217   structures, they can be problematic and inapplicable in some of the most interesting,

218   highly variable regions, particularly as comparisons expand across increasingly

219   disparate taxa. Similarly, surface points cannot always be applied to all structures, such

220   as the extremely narrow palatal region of snakes. Both curve and surface sliding

221   semilandmarks provide important and complementary information on shape variation

222   and our results demonstrate that both are improvements over analyses of landmarks

223   alone for characterizing complex morphologies.

224   This result is further demonstrated by examining patterns of variance across

225   landmarks and semilandmarks (Fig. 2). While the overall distribution of variance is

226   similar in both datasets, large areas of the cranium are unsampled in landmark-only

227   datasets, and thus some regions that are highly variable across taxa, such as the

228   maxillopalatine of caecilians, are inadequately represented by landmarks. Thus, high-

229   density configurations clearly contain important aspects of shape variation that is not

230   captured by landmark-only analyses.

231

The effect of Procrustes superimposition on analyses of modularity

233   In order to assess how Procrustes superimposition impacts covariance patterns

234   between landmarks and the ability to recover modular patterns from them, we

235   performed a controlled series of simulation experiments in which we varied the degree

236   of variability at each landmark, the direction of covariation, and the number of

237   landmarks. Each experiment is described in detail below.

238   Experimental samples were modelled by randomly perturbing landmarks around

239   a base configuration (or "*archetype*"; Fig. 3A) based on a multivariate normal covariance

240   matrix **V** that we varied systematically with each experiment (Fig. 3B). Each instance of

241   **V** was given two modules in which covariances among landmarks (and semilandmarks)

242   within modules was higher than between modules. The number of rows and columns

243   (landmark coordinates) in **V** and the magnitude of their covariances was varied to match

244   the conditions of each experiment. Residual variation was then simulated by post-

245   multiplying the Cholesky decomposition of **V** by a *kp* x *n* matrix of points drawn from *n*

246   univariate normal distributions with mean of 0 and variance *v*, where *k* is the number of

247   landmarks (and semilandmarks), *p* is the dimensionality of each landmarks (or

248   semilandmark), and *n* is the number of individuals in the sample. This multiplication

249   produces a matrix of *n* individuals with *kp* landmarks (and semilandmarks) with

250   covariance **V**. Finally, the residuals were added to the base configuration of landmarks

251   (and semilandmarks) to produce a sample of shapes (Fig. 3D). Each simulated dataset

252   consisted of 500 individual shapes unless otherwise noted.

253   Note that covariance between the x and y (and z) axes of a landmark produces a

254   scatter of variation that has a directional orientation. For example, if a landmark has

255   equal variances in both the x and y axes, any covariance between them will produce an

256   ellipse of points with a major axis at an angle of 45º. For convenience, all coordinates

257   were given the same variance, which produced this 45º angle in all landmarks (either in

258   a positive or negative direction). For experiments where a more directionally complex

259   covariance pattern was desired, individual scatters of simulated residual points were

260   rotated into new orientations (i.e., the ellipsoids in Figure 3B were pivoted around their

261   corresponding landmark into new orientations), which is equivalent to altering the

262   variances and covariances of their coordinates.

263   In each experiment, we assessed the effect of Procrustes superimposition on

264   recoverability of modules using two metrics: (1) we tested whether the original modular

265   pattern was significantly supported after Procrustes superimposition using the CR

266   coefficient randomization test (Adams 2016) and (2) we compared the modules

267   recovered from the original and Procrustes superimposed shapes using hierarchical

268   clustering analysis. The CR test determines whether ratio of covariation within and

269   between the original modules is strongly enough preserved to produce a statistically

270   significant correlation compared to randomized modules. CR values are high when

271   between module correlations are higher than within module correlations (i.e., when

272   modules are not distinct) and they decline toward 0 as modularity becomes stronger.

273   Significance is tested by randomizing landmarks between modules and comparing the

274   observed CR value with the distribution of randomized values (Adams 2016). The

275   hierarchical clustering analysis used Ward's minimum variance linkage algorithm on a *k*

276   *x k* covariance matrix using canonical correlations between landmarks (Goswami and

277   Polly 2010). This approach minimizes total within-cluster variance to cluster landmarks

278   and was used to determine whether the same organisation of traits (i.e. modules) was

279   recovered before and after Procrustes superimposition and whether that pattern

280   matched the modules constructed in **V**. Hereafter, we refer to the original simulated

281   shapes before Procrustes superimposition as "naturally superimposed", and we discuss

282   the assumptions and implications of that concept further below.   The number of

283   significant modules in each cluster was estimated by comparing the observed

284   eigenvalue structure to a null distribution derived from a Monte Carlo simulation using

285   the same base shape but with zero covariance with 100 iterations (see Goswami and

286   Polly 2010; Polly and Goswami 2010). All analyses were performed in *Mathematica*

287   (Wolfram Research, 2018) using the *Modularity for Mathematica* (v. 2.0) and *Geometric*

288   *Morphometrics for Mathematica* packages (Polly 2019; Polly and Goswami 2010).

289

Experiment 1: Direction of covariance

In this experiment, the direction of landmark covariance was systematically altered (Fig. 3E-G). A simple *archetype* of eight landmarks arranged in a rectangle with two modules of four landmarks symmetrically arranged to the left and right of the *archetype*'s centroid was used. Correlations between landmarks within each module was set at 0.8, except for the second test where one module was given completely invariant landmarks except for a small amount of uncorrelated noise. In the first test, the orientation of covariance in the left module was set at positive 45° with respect to the length of the archetypal rectangle and in the right module it was set at 135°, which is 90° to the first module (Fig. 3E). In the second test, the left module had four invariant landmarks and the right module was identical to the right module in the first test (Fig. 3F). In the third test of this experiment, the orientation of variation in both modules was such that each landmark had a positive covariance pointing away from its respective module's center (Fig. 3G).

In the first test in this experiment, Procrustes superimposition altered the covariance pattern so much that the original modules were unrecoverable. Despite having a strongly modular pattern that was easily recovered from the naturally superimposed data, the modules were not recovered from the Procrustes superimposed shapes. The pattern of covariance was strongly altered by Procrustes superimposition, which is seen visually in Figure 3E and indicated by their comparatively high CR value (CR = 1.27; P = 0.94). Note that the centroids of the original shapes are highly variable in their position, with an unconstrained scatter that is nearly as large as the scatter of points around any of the landmarks (Fig. 3E). The stability of the centroid point turns

313   out to be an important factor determining how much Procrustes superimposition alters

314   the covariance pattern of the landmarks.

315   The second test, in which one module consisted of invariant landmarks,

316   performed no better and arguably worse in terms of module recoverability (Fig. 3F).

317   The two modules were not recoverable even from the naturally superimposed data,

318   largely because the "invariant" module is not truly modular because its landmarks do not

319   covary. The dendrogram based on the naturally superimposed shapes recovered a

320   tight cluster between the four landmarks in the right module, but they were not

321   significantly distinguished from the landmarks of the left "module" based on the

322   eigenvalue variance randomization tests. Similarly, only one module was recovered

323   from the Procrustes superimposed data, but there was no hint of similarity between the

324   landmarks of the right module in the dendrogram. CR was also high and non-significant

325   (CR=1.14; P=0.30). The position of the centroid of the naturally superimposed shapes

326   was more constrained than in the first test, although it was still quite variable.

327   In the third test, in which the direction of variation was symmetrically radial in

328   each module instead of perfectly parallel, the true modular pattern was easily recovered

329   (Fig. 3G). Variability in the position of the centroid in the naturally superimposed shapes

330   was much less than in the previous two tests, and much smaller than the variability at

331   individual landmarks. The relative consistency of the position of the centroid is a result

332   of the symmetry of the landmark variability. Because the original centroids are close

333   together, changes in the overall pattern of covariance due to Procrustes superimposition

334   are small. The CR test indicated that the original modules were recoverable after

335   Procrustes superimposition (CR=0.51; P=0.00).

336   This experiment suggests that the symmetry (or lack thereof) in the directions of

337 covariance patterns within and between modules affects variability in position of the

338 centroid from one shape to the next and that the degree of variation in the position of

339 the centroid relative to variation in individual landmarks is a major determinant of how

340 much Procrustes superimposition, which recenters shapes on their centroids, alters the

341 covariance structure.

342 *Experiment 2: Magnitude of variance*

343   One possible interpretation of the first experiment is that the less variation there is in

344 shape, the more constrained will be the position of the centroid and the less the

345 covariance pattern will be altered by Procrustes superimposition. In the second

346 experiment, we therefore tested whether the magnitude of shape variation has an effect

347 on recoverability of modular patterns. It does not.

348   This experiment used the same directional covariance structure as in the first test of

349 the previous experiment (Fig. 3E) but systematically varied the amount of variance in

350 the landmark coordinates (Fig. 3H-J). The first test in Experiment 2 was stochastically

351 identical to the first test in Experiment 1 (CR=1.25; P=0.93). In the second and third

352 tests, the variance at each landmark was reduced to 80% and 60% respectively (and

353 the strength of covariance was maintained at 0.8). Even though variation in the position

354 of the centroid was progressively smaller in the second and third tests (Fig. 3I, J), the

355 CR coefficient remained approximately the same (CR=1.24 & 1.25; P=0.90 & 0.93) and

356 the original modules were not recovered from the Procrustes superimposed data.

357   Even though the centroid position was less variable in the second and third tests,

358   the effect of Procrustes superimposition on the covariance structure remained

359   approximately constant because the centroid remained just as variable with respect to

360   the variation at the individual landmarks. The translational and rotational components of

361   Procrustes superimposition therefore had a proportionally similar effect on the relative

362   positions of the landmarks (and therefore their covariance structure) regardless of the

363   absolute magnitude of shape variation. This experiment shows that it is not the

364   magnitude of shape variation *per se* that matters.

365   *Experiment 3: Number of landmarks*

366   The third experiment doubled and tripled the original number of landmarks to

367   determine whether additional landmarks help minimize the effect of Procrustes

368   superimposition (Fig. 4A-C). They do not (at least not without the contribution of other

369   factors, as explained below). The first test in this experiment (Fig. 4A) was

370   stochastically identical to that in Figure 3E (CR=1.28; P=0.96). In the second test, four

371   new landmarks were added to each module positioned one-quarter of the way toward

372   the respective center of the module (Fig. 4B). In the third test, four more landmarks

373   were added, these equidistant from the original four landmarks along the periphery of

374   each module (Fig. 4C). The direction of covariation of the new landmarks in each

375   module was identical to its original four.

376   The addition of landmarks had no substantial effect on variation in the position of

377   the centroid of the naturally superimposed shapes, and only minor improvements in the

378   CR test (CR=1.11 & CR=1.09; P=0.88, P=1.00) and offered no improvement in the the

379   recoverability of modules. Because the additional landmarks covary in the same

380   direction and with the same magnitude as the original landmarks, they do not constrain

381   the position of the centroid and are thus equally affected by the Procrustes

382   superimposition process. Therefore, the effects of Procrustes superimposition on

383   covariance structure are not increased by the addition of landmarks (or semilandmarks),

384   *contra* Cardini (2018), but neither are they decreased.

Experiment 4: Direction of Covariance II

386   The first three experiments indicate that Procrustes superimposition has a strong

387   effect on the covariance matrix, and thus recoverability of modules, when variation in

388   position of the centroid is only loosely constrained relative to variation in the individual

389   landmarks. Neither the absolute variability nor the number of the landmarks have an

390   effect, but the overall pattern of directionality of covariation in the landmarks does. The

391   effect of Procrustes superimposition was minimized in the third test of the first

392   experiment when directionality of variation was symmetric with respect to both the

393   center of each module and the centroid of the entire shape.

394   Next, we tested how random patterns of directional variation within and between

395   modules affect recoverability of modules (Fig. 4D-F). Variation in real biological

396   structures is much more directionally complex than any of the examples tested in the

397   first experiment (e.g., Zelditch *et al.* 1993). It is difficult to imagine a biological example

398   in which trait variation across a complex morphology is structured in entirely parallel or

399   perpendicular directions. Thus, in this experiment we randomly oriented the direction of

400   covariance at each landmark to produce a pattern that is not strictly symmetric as in the

401   third test of the first experiment, but which varies in a more complex, and arguably more

402   "biological", manner than any of the examples in the first experiment.

403   The first test of experiment 4 used parameters identical to the first in experiment 2

404   as a reference (Fig. 4D; CR=1.27; P=0.96), but in the second two tests (Fig. 4E, F) the

405   directions of variation at each landmark were randomly rotated by 0º to 360º. In both

406   cases, the effect was to dramatically constrain the position of the centroid with respect

407   to the variation in the landmarks, to improve recoverability as measured by CR

408   (CR=0.42 & 0.74; P=0.00 & 0.01), and to recover the original modular patterns

409   accurately. While Procrustes superimposition had a small effect on the covariance

410   matrix and the perceived closeness of relation between landmarks in each module, this

411   effect was minimal.

412   The results of the first experiment can now be reinterpreted in light of the fourth: it

413   is not symmetric shape variation that matters as much as the lack of systematically

414   directional variation. In both the first and second tests of the first experiment, the

415   direction of variation at all landmarks was somewhat parallel. In the first experiment all

416   of the landmarks shared half of their variation as a vertical component, whereas in the

417   second experiment all of the landmarks that varied shared their direction. The

418   symmetrical pattern in the third test of the first experiment performed no better than the

419   random patterns in the second and third tests of the fourth experiment. Regardless of

420   whether the landmark variation is directionally random or symmetrical, the effect is to

421   severely constrain variation in the position of the centroid relative to the landmarks, and

422   therefore to minimize the effects of Procrustes superimposition on the covariance

423   matrix.

19

Experiment 5: Direction of covariance and number of landmarks

   If the complexity of the directional variation matters, then more landmarks should

increase that complexity if their direction of variation is independent. We tested that

possibility in our fifth and final experiment (Fig. 4G-I). We used the same 8, 16, and 24

landmarks as in the third experiment, but this time randomly rotated the direction of

variation at each landmark. When the major axis of variation at each landmark is

oriented in a different direction, increasing the number of landmarks has a positive

effect on the recoverability of modules. As the number of landmarks increased, the CR

ratio declined (CR = 0.34 and P = 0.00 for k = 8, CR = 0.17 and P=0.00 for k= 16 and

CR=0.18 and P=0.00 for k=24). With 24 landmarks with randomly varying directionality,

Procrustes superimposition had little visible effect on the covariance pattern or on the

modularity dendrogram (Figure 4I).

Further considerations on centroids and natural superimpositions

   The original simulated shapes before Procrustes superimposition can be considered

to be in their "natural" superimposition, especially if the base shape has a centroid size

of one. The concept of "natural superimposition" warrants philosophical consideration.

It is a biologically vague idea, yet the crux of the issue of whether Procrustes

superimposition alters the "real" covariances between landmarks depends upon the

idea of a "natural superimposition". The strategy of the Procrustean paradigm in

geometric morphometrics is to remove so-called "nuisance" parameters of size,

translation, and rotation by translating landmarks (and semilandmarks) so that the

centroid of each shape is at the origin, scaling them to have centroid size of one, and

446   rotating them to minimize the sum-of-squared distance between shapes. Upon

447   completion of the superimposition, the new shape data are placed in a single

448   comparable coordinate system where their differences can be analysed, analogous to

449   mean-centering normal variables and standardizing them to unit variance. The strategy

450   we adopt here assumes that individuals are generated by some process (e.g.,

451   ontogenetic development) that produces variants on a general theme (our base

452   landmark configuration, which we refer to as the *archetype* after Richard Owen's notion

453   that vertebrate species were all variations on an underlying theme) with a covariance

454   structure **V** that arises from the generating process. Since our modelling procedure

455   (Fig. 3A) generates residual variation from a multivariate normal covariance distribution

456   with a mean of zero, the shapes are invariant with respect to translation and rotation;

457   and since the residuals are all added to the same *archetypal* configuration of landmarks

458   (and semilandmarks), they are also invariant in scale with respect to the process that

459   generated them.

460   Individual simulated shapes, however, do not have a centroid size of one, their

461   individual centroids are not aligned, they are not in optimal alignment, and their shapes

462   are not the same as the *archetype*. Figure 3C shows two simulated shapes along with

463   their centroids to illustrate this fact. Instead, having a centroid size of one, a centroid

464   centred at the origin, and an archetypal shape are properties of the mean of the

465   simulated shapes (Fig. 3D).   Thus, the simulated shapes are not aligned using

466   Procrustes superimposition, but they are in the optimal alignment with regard to the

467   process that generated them. This difference between the two alignments is the source

468   of Procrustes-induced covariance patterns. Accurately representing the natural

469   superimposition, and thus the processes generating shape variation, is a critical

470   concern in most analyses employing geometric morphometrics, and thus understanding

471   the cause of these deviations is an important theoretical and practical consideration.

472   The reason why the centroids are not perfectly aligned is because the generating

473   process used in these examples makes no explicit reference to the centroid. Instead,

474   the generating process produces random deviations from an archetypal configuration of

475   landmarks with a modular covariance pattern. Each deviation has its own centroid,

476   centroid size, and orientation relative to the archetype. One can imagine other

477   generating processes that do make reference to the centroid (or, at very least, to a

478   landmark that has an invariant position). For example, the development of the

479   tribosphenic molar involves a process of tissue growth that begins with the apex of a

480   particular tooth cusp (the protoconid) and via a cascade of molecular signalling and

481   folding produces additional cusps in a complex pattern around the original one (Jernvall

482   1995; Thesleff and Sahlberg 1996). One can therefore say that the natural alignment of

483   tribosphenic tooth shapes is invariant at the protoconid cusp tip with a variance and

484   covariance structure determined by the cascade of subsequent cusp formation. Polly

485   (2005) simulated tooth shapes using an analogous cascading process that started with

486   the protoconid landmark. But even in this example, the protoconid cusp is not

487   equivalent to the centroid, which varies in its relative position depending on the

488   arrangement of other cusp landmarks.

489   If there were a generating process that began with an object's centroid, such as

490   development of a radially symmetric structure like a coral polyp (c.f., Budd *et al.* 1994)

491   the "natural" and Procrustes superimpositions could be nearly identical once

492 standardized for size, rotation, and translation. But, as our experiments show, a

493 complex pattern in the direction of variation around landmarks with respect to one

494 another coupled with strong covariance has the effect of constraining the location of the

495 centroid, regardless of the generating process. The greater the complexity, the greater

496 the constraint on the centroid position, and the more similar the "natural" and Procrustes

497 superimpositions.

498   Presuming that real biological shapes have similar directional diversity of landmark

499 variation within modules as in our fifth experiment, our results suggest that Procrustes

500 superimposition is unlikely to interfere with the recoverability of modular patterns, even

501 when the number of landmarks is small. Properties that matter for recoverability of

502 modular patterns include: 1) variation in directional variation within and between

503 modules and 2) centroids whose "natural" position varies little in proportion to variation

504 in individual landmarks. Properties that do not matter for recoverability of modular

505 patterns include: 1) total number of landmarks (or semilandmarks) and 2) absolute

506 magnitude of shape variation.

507   Thus on the question of whether the use of sliding semilandmarks exacerbates the

508 effect of Procrustes superimposition on covariance structure (Cardini 2018), the results

509 of our third experiment suggest that adding landmarks neither improves nor inhibits the

510 recoverability of modules. The fact that the direction of variation in sliding

511 semilandmarks tends to be fairly uniform as a result of their fitting procedure (e.g.,

512 Perez *et al.* 2006) suggests that they will not improve recoverability to the same extent

513 as covarying landmarks (or non-sliding semilandmarks) whose direction varies with

514 respect to one another. However, sliding semilandmarks improve representation of

515   complex structures, such as surfaces, far beyond the abilities of landmarks, and thus

516   the increased complexity, and added variation in directionality of variation, will constrain

517   centroid variation, improve the Procrustes fit relative to the 'natural superimposition',

518   and thus increase the accuracy of recovering modules for biological structures.

519

Comparing analyses of integration with landmark and semilandmark datasets

521   In the above sections, we demonstrate that high-density semilandmark datasets add

522   important detail on morphology beyond that which is captured by Type I/II landmarks.

523   In addition, our simulations indicate that Procrustes superimposition does not mislead

524   analyses of integration in biologically realistic scenarios, i.e. those with complex

525   directions of variation sampled by geometric morphometric data, regardless of number

526   of landmarks or semilandmarks. Finally, we address the question of how using

527   semilandmarks in analyses of integration and modularity may change results and

528   interpretations of these quantities, compared to analyses based on landmarks alone.

529   Because semilandmarks and sliding semilandmarks are not independent of each other

530   due to their fitting procedure, there are expected effects on analyses of integration and

531   modularity. Specifically, adjacent semilandmarks and sliding semilandmarks will be

532   correlated because their placement is relative to each other, in addition to any biological

533   correlation amongst the structures they represent. The effect of this fitting may be to

534   exaggerate the correlations or covariance of proximal semilandmarks relative to those

535   farther away, which may increase the appearance of modularity across regions. On the

536   other hand, landmarks (and also curves based on element boundaries) may have the

537   opposite effect. Because Type I landmarks in a structure such as a skull will be largely

538   limited to sutures between elements, they may suffer from boundary bias, exaggerating

539   the apparent integration of those elements compared to aspects of their respective

540   morphologies that are not located at their point of juncture. It is important to recognize

541   that both approaches suffer from statistical artefacts due to the nature of the data

542   collection approach and may have opposing biases in reconstructing trait integration

543   and modularity. Thus, the comparison of results generated by these different

544   approaches is critical for identifying the magnitude and impact of their respective biases

545   and artefacts.

546   In two recent studies of variational or static (Marshall *et al.* 2019) and

547   evolutionary (Bardua *et al.* 2019b) integration and modularity in caecilian crania, we

548   conducted extensive analyses of integration across 16-17 cranial regions using 66

549   (*Idiocranium russeli*), 68 (*Boulengerula boulengeri)* or 53 (32 caecilian genera)

550   landmarks and 1363-1558 curve and surface sliding semilandmarks. These datasets

551   were analysed using Covariance Ratio (CR) analysis (Adams 2016) and a maximum

552   likelihood approach (Goswami and Finarelli 2016), with allometric and phylogenetic (for

553   the intergeneric analysis) corrections. In both studies, results were compared across

554   analyses of the full dataset and analyses of the landmark-only datasets. In the

555   intergeneric study of evolutionary modularity, both datasets significantly supported a

556   highly modular pattern (16 module model, full dataset CR = 0.59, p < 0.01; landmark-

557   only dataset CR = 0.88, p < 0.01). Despite supporting a modular pattern, the landmark-

558   only dataset returned a CR much closer to one, indicating relatively more integration

559   among modules. In particular, the major differences were increased integration of the

560   bones forming the cranial vault, which, in landmark-only analyses are defined entirely by

561   their sutures (mainly with each other), and reduced within-region integration in the

562   landmark-only analyses, as expected. (Bardua *et al.* 2019b; SI Fig. 2). A similar result

563   is observed in the intraspecific study of two species of caecilieans (Marshall *et al.* 2019),

564   with all analyses again significantly supporting a highly modular skull. For example,

565   covariance ratio analyses of the 17-module model for *Idiocranium russeli* were highly

566   significant for the full dataset before (CR = 0.621, p < 0.001) and after (CR = 0.519, p <

567   0.001) allometric correction and with the landmark-only dataset before (CR = 0.851, p <

568   0.001) and after allometric correction (CR = 0.738, p < 0.001). As before, the landmark-

569   only analyses returned CR values closer to one, suggesting more integration than the

570   analysis of the full dataset, and removing allometric effects resulted in reduced CR

571   values, supporting a more modular pattern. Despite this overall consistency across

572   datasets and analyses, examination of the pairwise CR values between regions, in

573   addition to the mean CR across the full cranium, suggests the allometry may have a

574   stronger influence on landmark-only analyses. For example, in the *Idiocranium russeli*

575   dataset, landmark-only analyses identify 49 out of 120 region pairs with CR values

576   greater than 0.9, with some exceeding a value of one (indicating integration). Following

577   removal of allometry, only 16 region pairs show CR values greater than 0.9, and the

578   overall pattern of integration across regions is congruent with the analysis of the full

579   dataset. Allometric correction did not have a similar effect on the analyses of the full

580   dataset. These results, while supporting that analyses are largely consistent across

581   datasets, suggest that allometry may have a stronger influence on recovered patterns of

582   integration in landmark-only datasets. If so, this effect may reflect the tendency for many

583 landmarks to be placed at element boundaries, resulting in a stronger signal of structure

584 size relative to the complexity of its shape, which the latter being better captured by

585 semilandmarks.

586

Conclusions

588 Capturing and quantifying morphology using high-resolution imaging has opened the

589 door to high-density morphometric data analysis with semilandmarks or

590 pseudolandmarks. Our analyses on both simulated and empirical datasets demonstrate

591 that semilandmarks provide far more comprehensive, as well as accurate,

592 characterizations of morphological variation than analysis of landmarks alone, which

593 suffer from limitations to points that can be identified repeatedly on specimens and often

594 leave large areas of complex structures entirely unsampled. However, these gains in

595 quantifying morphology raise questions about the biases that these datasets may bring,

596 in terms of quality of data, procedural artefacts, and ability to accurately recover

597 attributes such as trait integration. Here we demonstrate that some of the concerns with

598 geometric morphometric analysis of trait integration and modularity are unlikely to affect

599 analyses of complex structures, such as those encountered in biological specimens. We

600 also demonstrate that increasing landmark or semilandmark sampling alone does not

601 exacerbate issues with procedures such as Procrustes analysis. We further suggest

602 that analyses incorporating semilandmarks may be less influenced by boundary bias

603 and allometric effects, which may exaggerate degree of integration across regions in

604 landmark-only analyses, while analyses of sliding semilandmark may exaggerate within-

605 region integration and between-region modularity. It remains a continuing challenge to

606 develop methods that alleviate these effects. In doing so, we should prioritize improving

607 the representation of morphology, rather than limiting future studies to existing methods

608 that quantify complex structures with a small number of lengths or landmarks and leave

609 much of the available biological information unused (Collyer *et al.* 2014). Similarly, most

610 existing methods for the analysis of phenotypic integration and modularity are overly

611 simplistic and incapable of accurately conveying the complex hierarchy of relationships

612 across traits. Furthermore, most of these methods have not been developed or tested

613 for high-density datasets, which will certainly present new challenges as these datasets

614 become increasing common in studies of phenotypic integration and morphological

615 evolution. It is thus critical to remember that all methods have costs and benefits,

616 including both landmarks and semilandmarks. Nonetheless, the benefits of high-density

617 geometric morphometrics for more precisely representing morphology solves many

618 issues with reconstructing the evolution of complex structures across disparate taxa and

619 is a promising path forward for "Big Data" approaches to comparative morphology.

620

627 This work is the product of many conversations with many people over many years, far

628 too many to name, but especially Miriam Zelditch, Dean Adams, Chris Klingenberg,

629 Paul O'Higgins, Julien Clavel, Marcela Randau, and Andrea Cardini. Some (maybe all)

630 of them will disagree with parts of it, and some of them may disagree with all of it, but

631 hopefully all of them don't disagree with all of it. Thank you to SICB and the organisers

632 of the "Big Data" symposium and to the editors and reviewers who offered insightful

633 comments on this work. Most of all, thank you to the people sharing their 3D scans

634 freely online and making this such an exciting time to be a comparative morphologist.

635

636

637 **References**

638 Adams D, Rohlf F, Slice D. 2004. Geometric morphometrics: ten years of progress following the
639   'revolution'. Hystrix 71:5-16.
640 Adams DC. 2016. Evaluating modularity in morphometric data: challenges with the RV
641   coefficient and a new test measure. Methods in Ecology and Evolution 7:565-572.
642 Adams DC, Rohlf FJ, Slice DE. 2013. A field comes of age: geometric morphometrics in the
643   21st century. Hystrix 24:7-14.
644 Andjelković M, Tomović L, Ivanović A. 2017. Morphological integration of the kinetic skull in
645   Natrix snakes. Journal of Zoology 303:188–198.
646 Baab KL. 2013. The impact of superimposition choice in geometric morphometric approaches to
647   morphological integration. Journal of Human Evolution 65:689-692.
648 Bardua C, Felice RN, Watanabe A, Fabre AC, Goswami A. 2019a. A practical guide to surface
649   sliding semi-landmarks in morphometric analyses. Integrative and Organismal Biology in
650   press.
651 Bardua C, Wilkinson M, Gower DJ, Sherratt E, Goswami A. 2019b. Morphological evolution and
652   modularity of the caecilian skull. BMC Evolutionary Biology 19:30.
653 Bookstein F, Streissguth A, Sampson P, Connor P, Barr H. 2002. Corpus callosum shape and
654   neuropsychological deficits in adult males with heavy fetal alcohol exposure.
655   Neuroimage 15:233–251.
656 Bookstein FL. 1991. Morphometric tools for landmark data: geometry and biology. Cambridge:
657   Cambridge University Press.
658 Boyer DM, Puente J, Gladman JT, Glynn C, Mukherjee S, Yapuncich GS, Daubechies I. 2015.
659   A new fully automated approach for aligning and comparing shapes. The Anatomical
660   Record 298:249-276.

661 Bright JA, Marugan-Lobon J, Cobb SN, Rayfield EJ. 2016. The shapes of bird beaks are highly
662   controlled by nondietary factors. Proceedings of the National Academy of Sciences, USA
663   113:5352-5357.
664 Budd AF, Johnson KG, Potts DC. 1994. Recognizing morphospecies in colonial reef corals. I.
665   Landmark-based methods. Paleobiology 20:484-505.
666 Cardini A. 2018. Integration and modularity in Procrustes shape data: is there a risk of spurious
667   results? Evolutionary Biology 46:90-105.
668 Collyer ML, Sekora DJ, Adams DC. 2014. A method for analysis of phenotypic change for
669   phenotypes described by high-dimensional data. Heredity 115:357-365.
670 Davies TG, Rahman IA, Lautenschlager S, Cunningham JA, Asher RJ, Barrett PM, Bates KT,
671   Bengtson S, Roger BJ Benson DMB, José Braga, Jen A Bright, Leon PAM Claessens,
672   Philip G Cox, Xi-Ping Dong, Alistair R Evans, Peter L Falkingham, Matt Friedman,
673   Russell J Garwood, Anjali Goswami, John R Hutchinson, Nathan S Jeffery, Zerina
674   Johanson, Renaud Lebrun, Carlos Martínez-Pérez, Jesús Marugán-Lobón, Paul M
675   O'Higgins, Brian Metscher, Maëva Orliac, Timothy B Rowe, Martin Rücklin, Marcelo R
676   Sánchez-Villagra, Neil H Shubin, Selena Y Smith, J Matthias Starck, Chris Stringer,
677   Adam P Summers, Mark D Sutton, Stig A Walsh, Vera Weisbecker, Lawrence M Witmer,
678   Stephen Wroe, Zongjun Yin, Emily J Rayfield, Philip CJ Donoghue. 2017. Open data
679   and digital morphology. Proceedings of the Royal Society B-Biological Sciences
680   284:20170194.
681 Du TY, Tissandier S, Larsson HCE. 2018. Integration and modularity of teleostean pectoral fin
682   shape and its role in the diversification of acanthomorph fishes. Evolution 73:401-411.
683 Felice RN, Goswami A. 2018. Developmental origins of mosaic evolution in the avian cranium.
684   Proceedings of the National Academy of Sciences of the United States of America
685   115:555-560.
686 Gonzales PN, Barbeito-Andres J, D'Addona LA, Bernal V, Perez SI. 2016. Performance of semi
687   and fully automated approaches for registration of 3D surface coordinates in geometric
688   morphometric data. American Journal of Physical Anthropology 160:169-178.
689 Goswami A, Finarelli JA. 2016. EMMLi: a maximum likelihood apprach to the analysis of
690   modularity. Evolution 70:1622-1637.
691 Goswami A, Polly PD. 2010. Methods for studying morphological integration and modularity. In:
692   Alroy J, Hunt EG, editors. Quantitative Methods in Paleobiology. Boulder, CO: The
693   Paleontological Society. p. 213-243.
694 Gunz P, Mitteroecker P. 2013. Semilandmarks: a method for quantifying curves and surfaces.
695   Hystrix 24:103-109.
696 Gunz P, Mitteroecker P, Bookstein FL. 2005. Semilandmarks in three dimensions. In: Slice DE,
697   editor. Modern Morphometrics in Physical Anthropology. Developments in Primatology:
698   Progress and Prospects. Boston, MA: Springer. p. 73-98.
699 Jernvall J. 1995. Mammalian molar cusp patterns: developmental mechanisms of diversity. Acta
700   Zool. Fenn 198:1-61.
701 Klingenberg CP. 2009. Morphometric integration and modularity in configurations of landmarks:
702   tools for evaluating a prior hypotheses. Evolution & Development 11:405-421.
703 Klingenberg CP, Marugan-Lobon J. 2013. Evolutionary covariation in geometric morphometric
704   data: analyzing integration, modularity and allometry in a phylogenetic context.
705   Systematic Biology 62:591-610.
706 Kulemeyer C, Asbahr K, Gunz P, Frahnert S, Bairlein F. 2009. Functional morphology and
707   integration of corvid skulls - a 3D geometric morphometric approach. Frontiers in
708   Zoology 6:2.
709 Larouche O, Zelditch ML, Cloutier R. 2018. Modularity promotes morphological divergence in
710   ray-finned fishes. Scientific Reports 8:7278.

711 Lele S, Richtsmeier JT. 1990. Statistical models in morphometrics: are they realistic?
712　　　　　Systematic Zoology 39:60-69.
713 Marshall A, Bardua C, Gower DJ, Wilkinson M, Sherratt E, Goswami A. 2019. High-dimensional
714　　　　　3D morphometric analysis supports conserved static (intraspecific) modularity in
715　　　　　caecilian (Amphibia: Gymnophiona) crania. Biological Journal of the Linnean Society in
716　　　　　press
717 Marugán-Lobón J, Buscalioni Á. 2004. Geometric morphometrics in macroevolution:
718　　　　　morphological diversity of the skull in modern avian forms in contrast to some theropod
719　　　　　dinosaurs. In: Elewa AMT, editor. Morphometrics: applications in biology and
720　　　　　paleontology. Berlin: Springer. p. 157–73.
721 Marugán-Lobón J, Buscalioni ÁD. 2003. Disparity and geometry of the skull in Archosauria
722　　　　　(Reptilia: Diapsida). Biological Journal of the Linnean Society 80:67-88.
723 Mitteroecker P, Gunz P. 2009. Advances in Geometric Morphometrics. Evolutionary Biology
724　　　　　36(2):235-247.
725 Monteiro LR, Abe AS. 1997. Allometry and morphological integration in the skull of Tupinambis
726　　　　　merianae (Lacertilia: Teiidae). Amphibia-Reptilia 18:397-405.
727 Ollonen J, Da Silva FO, Mahlow K, Di-Poi N. 2018. Skull development, ossification pattern, and
728　　　　　adult shape in the emerging lizard model organism *Pogona vitticeps*: a comparative
729　　　　　analysis with other squamates. Frontiers in Physiology 9:1-26.
730 Olson EC, Miller RL. 1958. Morphological Integration. Chicago University of Chicago Press.
731 Parr W, LAB W, Wroe S, Colman N, Crowther M, M L. 2016. Cranial Shape and the Modularity
732　　　　　of Hybridization in Dingoes and Dogs; Hybridization Does Not Spell the End for Native
733　　　　　Morphology. Evolutionary Biology 43:171–87.
734 Perez SI, Bernal V, Gonzalez PN. 2006. Differences between sliding semi-landmark methods in
735　　　　　geometric morphometrics, with an application to human craniofacial and dental variation.
736　　　　　Journal of Anatomy 208:769-784.
737 Polly PD. 2005. Development and phenotypic correlations: the evolution of tooth shape in Sorex
738　　　　　araneus. Evolution & Development 7(1):29-41.
739 Polly PD. 2008. Developmental dynamics and G-matrices: Can morphometric spaces be used
740　　　　　to model phenotypic evolution? . Evolutionary Biology 35:83-96.
741 Polly PD. 2019. Geometric Morphometrics for Mathematica 12.3 ed.: Department of Earth and
742　　　　　Atmospheric Sciences, Indiana University: Bloomington, Indiana.
743　　　　　https://pollylab.indiana.edu/software/.
744 Polly PD, Goswami A. 2010. Modularity for Mathematica, Version 2.0. . Department of Earth
745　　　　　and Atmospheric Sciences, Indiana University: Bloomington, Indiana.
746　　　　　https://pollylab.indiana.edu/software/.
747 Pomidor BJ, Makedonska J, Slice DE. 2016. A landmark-free method for three-dimensional
748　　　　　shape analysis. PLoS ONE 11:e0150368.
749 Richtsmeier JT, Lele SR. 2001. An invariant approach to statistical analysis of shapes. Florida:
750　　　　　Chapman and Hall/CRC.
751 Rohlf F, Marcus L. 1993. A revolution in morphometrics. Trends in Ecology & Evolution 8:129-
752　　　　　132.
753 Rohlf FJ. 1990. Rotational fit (procrustes) methods. In: Rohlf FJ, Bookstein FL, editors.
754　　　　　Proceedings of the Michigan morphometrics workshop. Ann Arbor: University of
755　　　　　Michigan Museum of Zoology. p. 227-236.
756 Rohlf FJ, Slice DE. 1990. Extensions of the Procrustes method for the optimal superimposition
757　　　　　of landmarks. Systematic Zoology 39:40-59.
758 Sanger TJ, DL M, A A, JB L. 2012. Roles for modularity and constraint in the evolution of cranial
759　　　　　diversity among Anolis lizards. Evolution 66:1525-1542.
760 Stayton CT. 2005. Morphological evolution of the lizard skull: a geometric morphometrics
761　　　　　survey. J Morphol 263(1):47-59.

762 Thesleff I, Sahlberg C. 1996. Growth factors as inductive signals regulating tooth
763   morphogenesis. Sem. Cell Develop. Biol. 7:185-193.
764 Tokita M, Yano W, James H, Abzhanov A. 2017. Cranial shape evolution in adaptive radiations
765   of birds: comparative morphometrics of Darwin's finches and Hawaiian honeycreepers.
766   Philosophical Transactions of the Royal Society B: Biological Sciences 372:20150481.
767 Urošević A, Ljubisavljević K, Ivanović A. 2018. Multilevel assessment of the Lacertid lizard
768   cranial modularity. (June):. Journal of Zoological Systematics and Evolutionary Research
769   57:145-158.
770 Vitek NS, Manz CL, Gao T, Bloch JI, Strait SG, Boyer DM. 2017. Semi-supervised
771   determination of pseudocryptic morphotypes using observer-free characterizations of
772   anatomical alignment and shape. Ecology and Evolution 7:5041-5055.
773 Watanabe A. 2018. How many landmarks are enough to characterize shape and size variation?
774   PLoS ONE 13:e0198341.
775 Watanabe A, Fabre A-C, Felice RN, Maisano JA, Müller J, Herrel A, Goswami A. 2019.
776   Ecomorphological diversification in squamates from conserved pattern of cranial
777   integration. Proceedings of National Academy of Sciences, USA in press.
778 Webster M, Zelditch ML. 2011. Modularity of a Cambrian ptychoparioid trilobite cranidium.
779   Evolution & Development 13:96-109.
780 Zelditch M, Swiderski DL, Sheets HD, Fink WL. 2004. Geometric morphometrics for biologists: a
781   primer. Boston: Elsevier Academic Press.
782 Zelditch ML, Bookstein FL, Lundrigan BL. 1993. The ontogenetic complexity of developmental
783   constraints. Journal of Evolutionary Biology 6:621-641.
784 Zelditch ML, Wood AR, Swiderski DL. 2009. Building developmental integration into functional
785   systems: function-induced integration of mandibular shape. Evolutionary Biology 36:71-
786   87.

787

Figure Captions

789 **Figure 1**. Characterization of morphologically disparate taxa. The disparity of biological

790 shapes and presence and absence of homologous structures, as exemplified in this

791 sample of diapsids and amphibians (A) and the difficulty of locating discrete landmarks

792 in some taxa, such as the strongly sutured skulls of birds (B) present challenges for the

793 quantitative analysis of morphology. High-density semilandmarks (C) can capture the

794 morphology of complex regions with far more detail and allow for comparisons of

795 homologous structures across disparate taxa, resulting in (D) massive increases in

796 dataset size for studies of comparative morphology (Felice and Goswami 2018)(Felice

797   and Goswami 2018)(Felice and Goswami 2018)(Felice and Goswami 2018)(Felice and

798   Goswami 2018; Watanabe *et al*. 2019).

799   **Figure 2**. Landmark-only (A, D) and full landmark and semilandmark configurations (B,

800   E) and landmark sampling curves generated by LaSEC for C) the frontal bone of

801   caecilians and F) the supraoccipital of squamates. Colours in A, B, D, and E indicate

802   Procrustes variance at each landmark position, demonstrating that full and landmark-

803   only configurations produce similar overall patterns but that some areas of high or low

804   variance are entirely unsampled in landmark-only analyses. Sampling curve (C, F)

805   illustrate that 25–35 landmarks and semilandmarks are required to confidently and

806   robustly characterize the shape variation in these individual bones.

807   **Figure 3**. Simulation experiments 1 and 2 of the effect of Procrustes superimposition

808   on covariance patterns and recovery of biological modules. Starting with a base

809   archetype (A), we perturbed variances and covariances (B) in each experiment, with

810   resultant effects on shape centroids (C), to generate a sample of "naturally

811   superimposed" shapes (D), which are then subjected to Procrustes superimposition. In

812   experiment 1, we test the effect of direction of covariance, with covariances of two

813   modules set at 90° to each other (E), one module of invariant landmarks (F), and both

814   modules with covariances oriented away from their respective centroids (G). In

815   experiment 2, we vary the magnitude of variance, with variances initially identical to that

816   of experiment 1 (H), and then reduced to 80% (I) and 60% (J). For each experiment,

817   landmark configurations are shown on the left, and clusters of recovered modules are

818   shown on the right.

819   **Figure 4**. Simulation experiments 3 – 5 of the effect of Procrustes superimposition on

820   covariance patterns and recovery of biological modules. In experiment 3, we increase

821   landmark numbers from the eight landmarks of experiment 1 (A), to 16 landmarks (B),

822   and 24 landmarks (C). In experiment 4, we vary the directionality of landmarks, from

823   the symmetric variation of experiment 1 (D) to random directions of variation (E, F).

824   Finally, in experiment 5, we combine the effects of experiments 3 and 4, by randomly

825   rotating landmarks for the initial set of 8 landmarks (G), and then 16 landmarks (H) and

826   24 landmarks (I). For each experiment, landmark configurations are shown on the left,

827   and clusters of recovered modules are shown on the right.

828

829

830

831

833 **Tables**

834

835 Table 1. Results from performing LaSEC with 1000 iterations on individual cranial

836 partitions of extant caecilian datasets. Values for Fit = 0.9, 0.95, and 0.99 denote the

837 median number of randomly subsampled landmarks degree of fit (0 to 1) of randomly

838 subsampled landmark configurations and fixed-only datasets to the respective full high-

839 dimensional coordinate data. Separate analysis of landmarks + curve sliding

840 semilandmarks was not conducted for caecilians, as curves for some regions (e.g.,

841 maxilla) were not homologous and removed prior to analyses. For details and definitions

842 of cranial regions, see Bardua *et al*. (2019b).

	# landmarks	# landmarks + semilandmarks	Fit = 0.90	Fit = 0.95	Fit = 0.99	Fit of landmark-only dataset
Basisphenoid	4	155	15	25	69	0.583
Frontal	4	125	13	21	61	0.617
Jaw joint	3	50	13	19	37	0.306
Maxillopalatine (interdental shelf)	4	110	13	19	52	0.782
Maxillopalatine (lateral surface)	3	134	14	23	64	0.238

| Maxillopalatine (palatal surface)	5	75	13	19	44	0.602
Nasopremaxilla (dorsal surface)	7	148	13	21	61	0.684
Nasopremaxilla (palatal surface)	3	59	8	12	29	0.770
Occipital condyle	2	34	11	15	27	NA (only two landmarks)
Occipital region	5	153	16	27	73	0.605
Parietal	3	126	11	18	51	0.361
Pterygoid	-	50	7	10	24	NA
Quadrate (lateral surface)	2	57	12	18	38	NA (only two landmarks)
Squamosal	4	104	15	25	61	0.574
Stapes	-	20	10	12	17	NA
Vomer	3	69	12	18	41	0.538

843

844

845   Table 2. Results from performing LaSEC with 1000 iterations on individual cranial

846   partitions of extant squamate datasets. Values for Fit = 0.9, 0.95, and 0.99 denote the

847   median number of randomly subsampled landmarks required for respective degree of fit

848   of randomly subsampled landmark configurations to the respective full (landmark +

849   curve and surface sliding semilandmark) dataset. Fit of Landmark-only and landmark +

850   curve sliding semilandmark datasets compared to full dataset is also provided for

851   comparison, demonstrating that the addition of curve sliding semilandmarks alone

852   greatly improves representation of shape over landmark-only analyses (although see

853   discussion regarding issues with curves for some highly-variable structures in the

854   caecilian skull). The occipital condyle, pterygoid, and palatine are not listed as they lack

855   either unique landmarks or surface sliding semilandmarks for some taxa. For details

856   definitions of cranial regions, see Watanabe *et al.* (2019).

Squamates	# LMs	# curve sLMs	# surface sLMs	Fit = 0.90	Fit = 0.95	Fit = 0.99	Fit of landmark-only dataset	Fit of landmark + curve dataset
Premaxilla	4	35	39	15	23	49	0.713	0.981
Nasal	4	40	42	15	25	54	0.664	0.977
Maxilla	5	65	92	16	27	74	0.696	0.913
Jugal	3	60	31	13	20	51	0.645	0.962
Frontal	4	40	86	14	25	66	0.721	0.993

| Parietal | 4 | 60 | 34 | 16 | 28 | 64 | 0.647 | 0.987 |
Squamosal	3	30	19	17	25	43	0.452	0.993
Jaw joint	4	20	18	20	27	38	0.484	0.999
Supraoccipital	5	60	67	30	55	90	0.597	0.979
Occipital condyle	-	15	22	22	27	34	N/A	0.988
Basioccipital	4	60	58	14	26	66	0.805	0.982

857

38
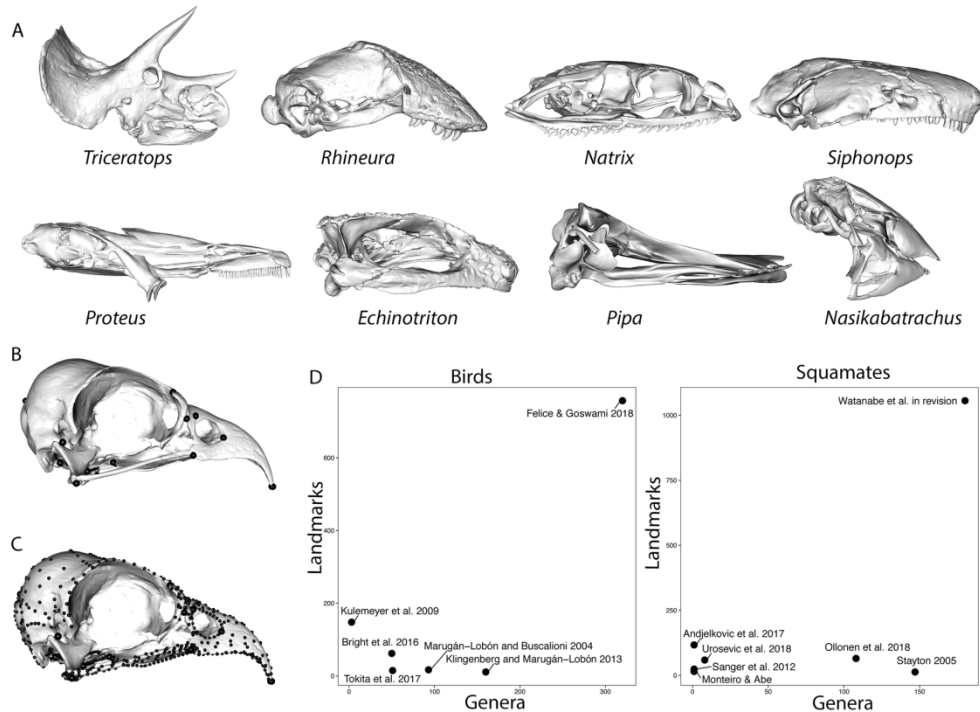
Characterization of morphologically disparate taxa. The disparity of biological shapes and presence and absence of homologous structures, as exemplified in this sample of diapsids and amphibians (A) and the difficulty of locating discrete landmarks in some taxa, such as the strongly sutured skulls of birds (B) present challenges for the quantitative analysis of morphology. High-density semilandmarks (C) can capture the morphology of complex regions with far more detail and allow for comparisons of homologous structures across disparate taxa, resulting in (D) massive increases in dataset size for studies of comparative morphology (Felice and Goswami 2018; Watanabe et al. in press).

160x117mm (300 x 300 DPI)
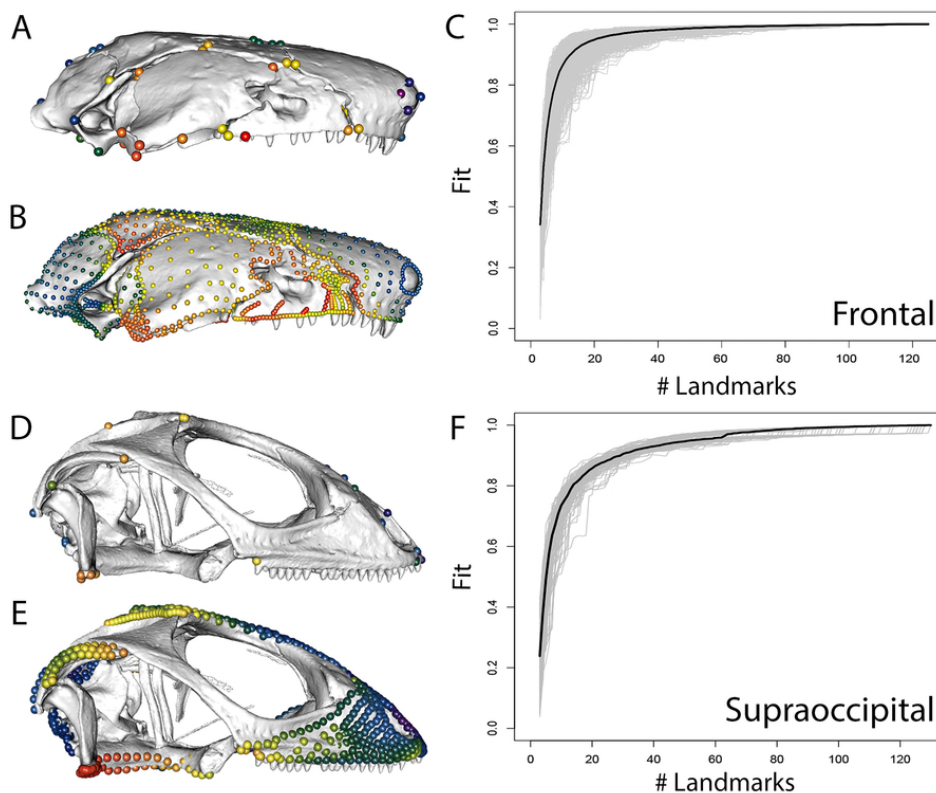
. Landmark-only (A, D) and full landmark and semilandmark configurations (B, E) and landmark sampling curves generated by LaSEC for C) the frontal bone of caecilians and F) the supraoccipital of squamates. Colours in A, B, D, and E indicate Procrustes variance at each landmark position, demonstrating that full and landmark-only configurations produce similar overall patterns but that some areas of high or low variance are entirely unsampled in landmark-only analyses. Sampling curve (C, F) illustrate that 25–35 landmarks and semilandmarks are required to confidently and robustly characterize the shape variation in these individual bones.

80x67mm (300 x 300 DPI)
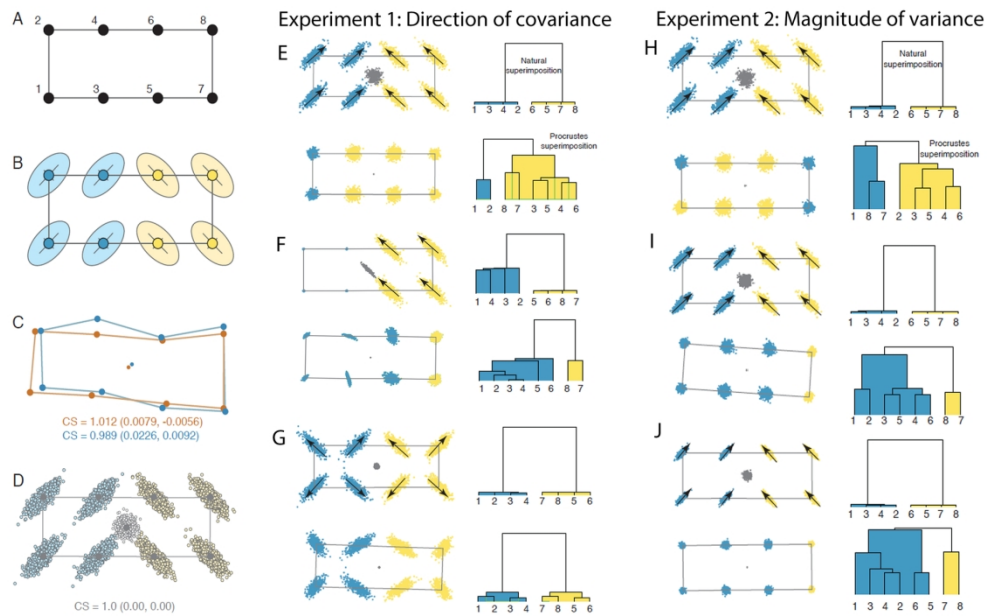
Simulation experiments 1 and 2 of the effect of Procrustes superimposition on covariance patterns and recovery of biological modules. Starting with a base archetype (A), we perturbed variances and covariances (B) in each experiment, with resultant effects on shape centroids (C), to generate a sample of "naturally superimposed" shapes (D), which are then subjected to Procrustes superimposition. In experiment 1, we test the effect of direction of covariance, with covariances of two modules set at 90° to each other (E), one module of invariant landmarks (F), and both modules with covariances oriented away from their respective centroids (G). In experiment 2, we vary the magnitude of variance, with variances initially identical to that of experiment 1 (H), and then reduced to 80% (I) and 60% (J). For each experiment, landmark configurations are shown on the left, and clusters of recovered modules are shown on the right.

160x103mm (300 x 300 DPI)
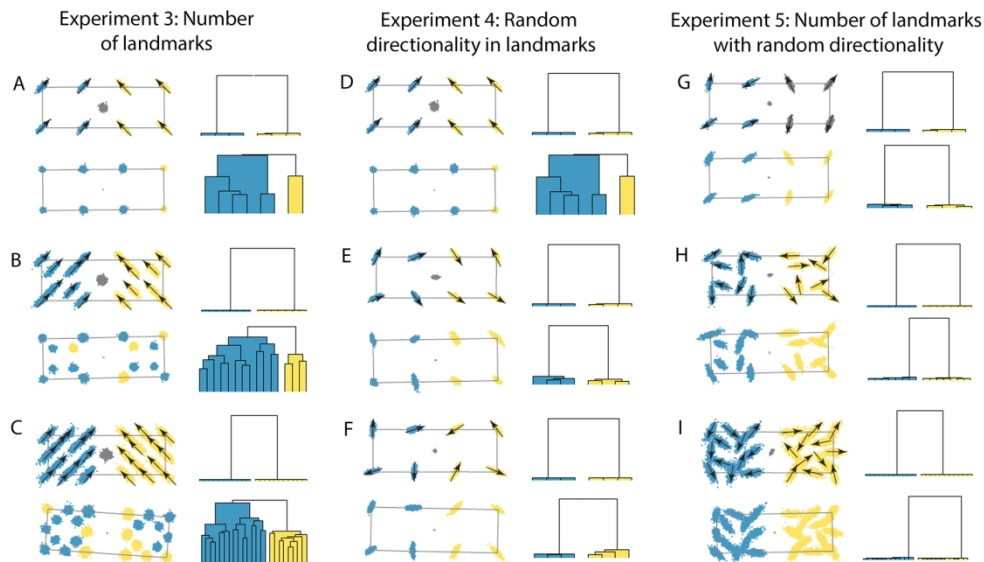
Simulation experiments 3 – 5 of the effect of Procrustes superimposition on covariance patterns and recovery of biological modules. In experiment 3, we increase landmark numbers from the eight landmarks of experiment 1 (A), to 16 landmarks (B), and 24 landmarks (C). In experiment 4, we vary the directionality of landmarks, from the symmetric variation of experiment 1 (D) to random directions of variation (E, F). Finally, in experiment 5, we combine the effects of experiments 3 and 4, by randomly rotating landmarks for the initial set of 8 landmarks (G), and then 16 landmarks (H) and 24 landmarks (I). For each experiment, landmark configurations are shown on the left, and clusters of recovered modules are shown on the right.

160x95mm (300 x 300 DPI)