Published by Associazione Teriologica Italiana
Hystrix, the Italian Journal of Mammalogy
Online first – 2017
OPEN ACCESS
Available online at:
http://www.italian-journal-of-mammalogy.it/article/view/12139/pdf
doi:10.4404/hystrix-28.2-12139

Research Article

Is there a "Wainer's rule"? Testing which sex varies most as an example analysis using GueSDat, the free Guenon Skull Database

Andrea Cardini[1,2,*], Sarah Elton[3]

[1] Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Via Campi 103, 41125 Modena, Italy
[2] Centre for Forensic Science, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
[3] Department of Anthropology, Durham University, Durham, DH1 3LE, UK

Keywords:
allometry
anatomical landmarks
bootstrap
geometric morphometrics
Procrustes
sampling error
shape
size

Article history:
Received: 4 October 2016
Accepted: 3 February 2017

Acknowledgements
This article is dedicated to the memory of Richard W. Thorington Jr. (Thor) 1937–2017. Thor has been an outstanding mammalogist and the most extraordinary man. He was unstoppable: when snow covered Washington DC, and most people could not go to work, he said that up to 10 cm he could still make it from his home in Bethesda to the Natural History Museum, and he was on a wheelchair. How much he has achieved, despite a most disabling disease, is an example for us all. We'll greatly miss him as a scientist and friend.

Many thanks to Howard Wainer, for inspiring the analysis of the magnitude of sex variances and for his kind and helpful replies to AC's questions about his original publication. Our gratitude also to R.K. Richardson, A. Pearson, A.P. Moczek, N. MacLeod, A. Pilastro and A. Palaoro for their fundamental help with references, greatly aiding us with the background to the study. We are truly grateful to K.B. Armitage, whom we asked with a very short notice to read the manuscript before the submission and whose supportive and experienced feedback was, as always, of incredible aid to finalize and improve the paper. We are also very grateful to Russell Hill and the Durham Primatology Group for their suggestions. Finally, we are in debt to the Associate Editor Paolo Colangelo, as well as to Pasquale Raia and an anonymous referee, for their most careful and helpful reviews, which greatly improved the paper. The data collection of the GueSDat by AC was funded by a grant to SE from the Leverhulme Trust (F/00128/T).

Abstract

The distinguished statistician Howard Wainer claimed that larger phenotypic variance in males might be a general occurrence in mammals. We called this putative pattern "Wainer's rule" and employed a dataset of more than 1300 specimens, each measured using 86 anatomical landmarks on skulls, to test this hypothesis using size and shape data in a group of Old World monkeys, the guenons. Our study is one example of an application that takes advantage of this large set of data (named "GueSDat"), made freely available to the research community. The analysis showed that large samples are crucial to estimate variances, and, in this respect, estimates of univariate size may require even larger samples than estimates of the magnitude of shape variance. Despite limited statistical power in species with smaller samples, results consistently suggest larger variance in male skull size but not in shape. Size could be more plastic and thus respond more directly to the environment. As males are larger than females, the costs of becoming bigger can be sustained only when conditions are optimal, thus making size strongly condition dependant and therefore more variable in the bigger sex. However, it is not only overall size and shape that may behave differently in terms of whether they follow "Wainer's rule": preliminary analyses suggest that, as in insects, different traits (e.g., different cranial regions) may vary in how similar or different their phenotypic variance is. The example study shows the potentially wide applications of data in GueSDat and suggests that, besides the most common comparison of mean differences in females and males, the study of differences between sexes in phenotypic variance offers a promising avenue for future research in mammals.

Indeed, as exemplified by our work, testing 'Wainer's rule' in mammals and other animals could become an active field of investigation in a variety of disciplines (from morphological to behavioural studies), and one that will hopefully elucidate whether this trend might be so common to be considered as a "rule" in evolutionary biology.

Introduction

The distinguished statistician Howard Wainer claimed that larger phenotypic variance in males might be a general occurrence in mammals, due to selection caused by male–male competition (Wainer, 2007, 255): *"Why was our genetic structure built to yield greater variation among males than females? And not just among humans, but virtually all mammals. … In most mammalian species … essentially all adult females reproduce, whereas only a small proportion of males do … One way to increase the likelihood of offspring being selected to reproduce is to have large variance among them. Thus evolutionary pressure would reward larger variation for males relative to females."* For the sake of brevity, we have, for the first time, coined this statement "Wainer's rule". We thus evaluate empirically whether this could be a well consolidated rule, such as Bergmann's or Allen's rule of morphological variation in relation to temperature (Gaston et. al, 2008), that should form the basis of significant future research[1]. In previous work on African monkeys (Cardini and Elton, 2008a), we noted that within-species skull size variation was greater in males than females, possibly because the longer growth period in males also resulted in a greater number of more extreme phenotypes. We also observed (Cardini et al., 2007) that male skull size was probably influenced more than female skull size by environmental factors. This finding is consistent with other work on baboons whereby female body mass seems less responsive than male mass to extrinsic environmental pressures, possibly because male growth, unlike that of females, is not truncated by reproduction and hence is influenced for longer by factors such as habitat productivity (Dunbar, 1990). Interestingly, a similar (albeit reversed) pattern has been found in many insects, most of which have female-biased sexual dimorphism. In such instances, and especially when females are much bigger than males, it appears that female body size is more sensitive than male size to environmental variation (Teder and Tammaru, 2005). It has also been suggested that secondary sexual characters may be more variable in the larger sex, at least when that is the male: it is assumed that the variation inherent in male ornamentation under sexual selection pressure has a cost, and that expression of the trait will depend on condition, creating high levels of variability, whereas the homologous characters in females are not shaped by

*Corresponding author
Email address: alcardini@gmail.com, cardini@unimo.it (Andrea Cardini)

[1] Wainer himself (*pers. comm.*) acknowledges that more evidence is needed to be confident on how widespread his "rule" might be, and, following Stigler's law of eponymy (Stigler, 1980), he even suspects that the "rule" may have already been in the literature, perhaps in a less explicit form.

sexual selection nor condition-dependent, and are hence more stable (Bonduriansky, 2007a).

Notwithstanding the precise explanations for and mechanisms by which greater phenotypic variation might occur, several other studies, across diverse taxonomic groups, have also indicated that one sex, often males, varies more than the other, even if variance homogeneity (i.e., homoscedasticity) has not been tested explicitly. Examples of greater male variance include: most cranial dimensions of Bornean orangutans (Leutenegger and Masterson, 1989); *Vipera berus* adders, where females are larger than males but males show larger differences in body size among populations (Forsman, 1991); sockeye salmon populations, where males are bigger and vary more in body length than females (Quinn and Foote, 1994); and *Polistes* wasps, where males show more variation than females in external morphology, although a similar pattern is probably not seen in *Vespa* (Eickwort, 1969 and references therein). Horned beetles have much greater variances in males than females, except in species where females are larger than males (Moczek, *pers. comm.*, and Moczek, 2006, also see Johns et al., 2014), and rock crabs also show multivariate heteroscedasticity, but we could not find a clear indication of which sex is more variable (Campbell and Mahon, 1974). Greater male variability, at least in some traits, has also been suggested in insects where males are the larger sex (Johns et al., 2014, and references therein), which may act as an honest condition-dependent signal (e.g., in relation to resource availability and competitive ability), used by females to select their mate and by males to assess or defeat their rivals. In contrast, there seems to be no sex-specific differences in variance in great tits (see data in Przybylo, 1995), human crania (Gonzalez et. al, 2011), the adult human pelvis (Tague, 1989; Bierry et al., 2010), and possibly (although it was not tested explicitly) human stature and body mass, at least in a British cohort (Power et al., 1997). In addition, teeth of archaeological samples of European migrants to Australia showed homoscedasticity for most linear measurements (Adler and Donlon, 2010).

Here, we examine "Wainer's rule", which refers specifically to increased phenotypic variance in male mammals, but may also be applicable across a much wider range of taxonomic groups. Very generally, a phenotype is the whole variety of biological traits above the level of gene and, although the term literally refers to the way an organism looks, it is not limited to morphological components and can be used to refer to, among many others, biochemical and physiological characters, as well as all aspects of behaviour, including culture and built structures, such as beavers' dams. We are especially interested in examining sex-based phenotypic variance from a morphological perspective, as this has received relatively little attention in mammalogy. Although sexual dimorphism has been the subject of many studies in primates (Cardini and Elton, 2008c, and references therein), the main focus has generally been on mean differences between sexes, and comparisons of the amount of phenotypic variation within sexes have seldom been assessed in detail.

In a similar fashion to many other "rules" in evolutionary biology, the one implied by Wainer's 2007 statement is likely to have exceptions and thus may be framed better as a dominant trend (i.e., the most frequent pattern found in a lineage). Although a strong test with robust results will require representatives of all main mammalian lineages, as well as a variety of phenotypic traits, as a first step towards exploring "Wainer's rule" empirically, we use skull measurements from a large Old World monkey sample, including all genera belonging to the African monkey tribe Cercopithecini, commonly called guenons, as well as two "outgroup" species. The outgroups are *Cercocebus atys*, a member of the same subfamily as the guenons (i.e., the Cercopithecinae), and *Colobus guereza*, a representative of the other cercopithecid subfamily (i.e., the Colobinae) (Grubb et al., 2003). However, for brevity, as the majority of the 1315 adult wild-caught individual monkeys are guenons, we will refer to this dataset simply as the "guenon skull database", or GueSDat. The dataset, consisting of Cartesian coordinates of 86 3D anatomical landmarks on the left side of the cranium and mandible, has been previously employed in a number of studies (e.g., Cardini and Elton, 2007, 2008a,b,c). As supplementary information accompanying this paper, it is now made freely available to all researchers interested in morphometrics and morphological evolution. The landmark data can be analysed using geometric morphometric methods (Adams et al., 2013; Cardini, 2013) or traditional morphometrics (Marcus, 1990), in which case the landmark data must simply be converted into linear distances prior to selecting the variables of interest among the >3600 resulting interlandmark measurements.

In this paper, we have two specific scientific objectives. The first is to test whether the magnitude of variance in skull size and shape is similar in females and males within the GueSDat species, and hence to undertake a preliminary investigation of "Wainer's Rule". As sample sizes are extremely heterogeneous across the different GueSDat species, the second scientific objective is to explore the sensitivity of the estimates of variance to sampling error in the two largest samples (N=146 and N=396). Sampling error is often neglected in morphometrics, but its effect on parameter estimates has been shown to be crucial even in relatively large samples (Cardini and Elton, 2007; Cardini et al., 2015, and references therein). By using bootstraps and randomized samples, we will show how critical it is to consider sampling error even in simple estimates such as the magnitudes of variance in size and shape, and also suggest that these two components of morphological variation may be differentially affected by sampling.

Materials and methods

General information about GueSDat

The list of landmarks included in GueSDat with their definitions is given in Tab. 1, with the left side configuration shown in Fig. 1. The 86 landmark configuration can be easily split into subsets (cranium, mandible, specific regions of the cranium, and so on) and used, for example, in modularity analyses (Cardini and Elton, 2008a; Klingenberg, 2013a). The species composition of GueSDat is detailed in Tab. 2. The classification largely follows Grubb et al. (2003) and corresponds to that reported in museum catalogues at the time of data collection (2004–2005). Sample sizes in GueSDat are very heterogeneous, generally slightly male-biased (as more male than female specimens tend to be represented in museum collections) and range from just one (*Cercopithecus solatus*) to almost 400 (*C. aethiops*) individuals.

A number of grouping variables, or classifiers, is provided in the GueSDat: an identifier, called "list" (i.e., a number corresponding to the original list of specimens built during the data collection); genus,

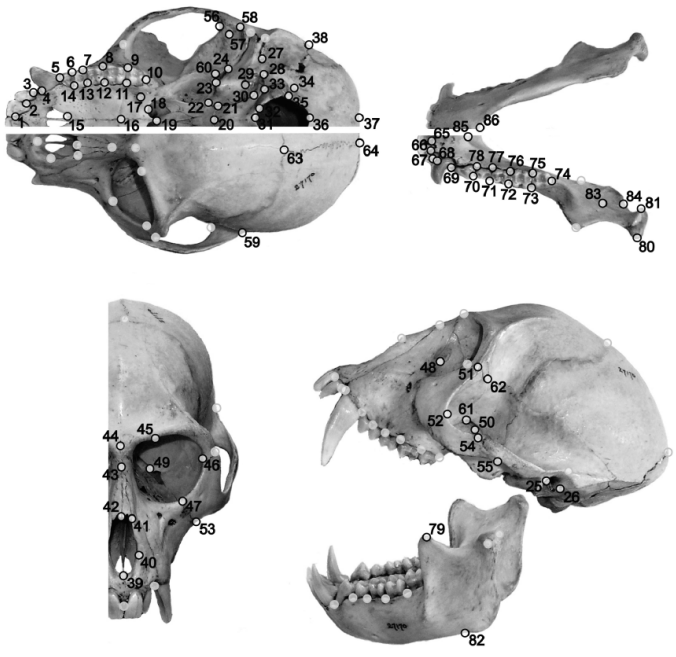

Figure 1 – Original left side 86 anatomical landmark configuration (modified from Cardini and Elton, 2007).

Table 1 – Definition and numbering of landmarks (L). The terms "anterior" and "posterior" are used with reference to Fig. 1. Landmarks 65 to 86 are on the mandible. Landmarks in parentheses are the "right side" ones, reconstructed by mirror-reflection of the left landmarks.

L	Definition
1	prosthion: antero-inferior point on projection of premaxilla between central incisors.
2 (87)	prosthion2: antero-inferiormost point on premaxilla, equivalent to prosthion but between central and lateral incisors.
3 (88)	posteriormost point of lateral incisor alveolus.
4 (89)	anteriormost point of canine alveolus.
5 (90)	mesial P3: most mesial point on P3 alveolus, projected onto alveolar margin.
6–9 (91–94)	contact points between adjacent premolars/molars, projected labially onto alveolar margin.
10 (95)	posterior midpoint onto alveolar margin of M3.
11–14 (96–99)	contact points between adjacent premolars/molars, projected lingually onto alveolar margin.
15	posteriormost point of incisive foramen.
16	meeting point of maxilla and palatine along midline.
17 (100)	greater palatine foramen.
18 (101)	point of maximum curvature on the posterior edge of the palatine.
19	tip of posterior nasal spine.
20	meeting point between the basisphenoid and basioccipital along midline.
21 (102)	meeting point between the basisphenoid, basioccipital and petrous part of temporal bone.
22 (103)	most medial point on the petrous part of temporal bone.
23 (104)	most medial point of the foramen lacerum.
24 (105)	meeting point of petrous part of temporal bone, alisphenoid and base of zygomatic process of temporal bone.
25–26 (106–107)	anterior and posterior tip of the external auditory meatus.
27 (108)	stylomastoid foramen.
28, 30 (109, 111)	distal and medial extremities of jugular foramen.
29 (110)	carotid foramen.
31	basion: anteriormost point of foramen magnum.
32, 35 (112, 115)	anterior and posterior extremities of occipital condyle along margin of foramen magnum.
33 (113)	hypoglossal canal.
34 (114)	center of condylar fossa.
36	opisthion: posteriormost point of foramen magnum.
37	inion: most posterior point of the cranium.
38 (116)	most lateral meeting point of mastoid part of temporal bone and supraoccipita. l
39	nasospinale: inferiormost midline point of piriform aperture.
40 (117)	point corresponding to largest width of piriform aperture.
41 (118)	meeting point of nasal and premaxilla on margin of piriform aperture.
42	rhinion: most anterior midline point on nasals.
43	nasion: midline point on fronto-nasal suture.
44	glabella: most forward projecting midline point of frontals at the level of the supraorbital ridges.
45 (119)	supraorbital notch.
46 (120)	frontomalare orbitale: where frontozygomatic suture crosses inner orbital rim.
47 (121)	zygo-max superior: antero-superior point of zygomaticomaxillary suture taken at orbit rim.
48 (122)	center of nasolacrimal foramen (fossa for lacrimal duct).
49 (123)	center of optic foramen.
50 (124)	uppermost posterior point of maxilla (visibile through pterygomaxillary fissure).
51 (125)	frontomalare temporale: where frontozygomatic suture crosses lateral edge of zygoma.
52 (126)	maximum curvature of anterior upper margin of zygomatic arch.
53 (127)	zygo-max inferior: antero-inferior point of zygomaticomaxillary suture.
54 (128)	zygo-temp superior: superior point of zygomaticotemporal suture on lateral face of zygomatic arch.
55 (129)	zygo-temp inferior: infero-lateral point of zygomaticotemporal suture on lateral face of zygomatic arch.
56 (130)	posteriormost point on curvature of anterior margin of zygomatic process of temporal bone.
57 (131)	articular tubercule.
58 (132)	distalmost point on post-glenoid process.
59 (133)	posteriormost point of zygomatic process of temporal bone.
60 (134)	foramen ovale (posterior inferior margin of pterygoid plate).
61 (135)	meeting point of zygomatic arch and alisphenoid on superior margin of pterygomaxillary fissure./
62 (136)	meeting point of zygomatic arch, alisphenoid and frontal bone.
63	bregma: junction of coronal and sagittal sutures.
64	lambda: junction of sagittal and lamboid sutures.
65	antero-superior point of mandible between central incisors.
66 (137)	antero-superior point of mandible between lateral incisors.
67 (138)	posteriormost point of lateral incisor alveolus.
68 (139)	anteriormost point of canine alveolus.
69 (140)	mesial P3: most mesial point on P3 alveolus, projected onto alveolar margin.
70–73 (141–144)	contact points between adjacent premolars/molars, projected labially onto alveolar margin.
74 (145)	posterior midpoint onto alveolar margin of M3.
75–78 (146–149)	contact points between adjacent premolars/molars, projected lingually onto alveolar margin.
79 (150)	superior tip of coronoid process.
80–81 (151–152)	most lateral and most medial points on mandible condylar surfaces.
82 (153)	anteriormost point on roughening for attachment of masseter on inferior margin of the angle of mandible.
83 (154)	mandibular foramen.
84 (155)	posteriormost point on superior area of insertion of medial pterygoid.
85	region of insertion of genioglossus muscles (midline posteriormost point on upper "ridge behind incisors"').
86	region of insertion of geniohyoid muscles (midline posteriormost point on lower "ridge behind incisors").

species, and sex; the museum acronym and the corresponding catalogue number of each individual; the original number of missing landmarks (between one to seven per specimen in 11% of the total sample, estimated by mean substitution, as detailed in Cardini and Elton, 2008a); a variable indicating if a specimen is suspected of being an outlier for shape, size or both (suspected outliers are those excluded from analyses in the current study); and also arbitrary numerical codes for each species (e.g., *Allenopithecus nigroviridis*=1, *C. atys*=101 etc.), and the two sexes (females=0, males=1), as well as variables to match data with a phylogeny (see below) and to rapidly select bigger samples (within sex N≥5). The identifier and the numerical codes for species and sexes may be ignored, and are in a sense redundant, but convey the most important information in a compact way, which is useful in some file formats commonly used in GMM, such as the `nts` format (Rohlf, 2015). Also, the sex dummy classifier can be imported easily as a covariate in MorphoJ (Klingenberg, 2011) and used to test sex differences and estimate how much variance they explain using a regression approach and permutations (Viscosi and Cardini, 2011). The classifier to match the data with a phylogeny allows selection of species for comparative analyses using phylogenetic independent contrasts in MorphoJ (see below).

Analysis of asymmetries (Klingenberg et al., 2002) was not an aim of the original project for which the GueSDat data were collected, so only the left side of the skull was measured. This allowed the sample size to be increased by approximately 40% compared to measuring both sides (Cardini, 2016). However, as recently shown (Cardini, 2016, 2017), using just one side of a structure with object symmetry can make size and especially shape estimates slightly inaccurate. A simple operation, which generally mitigates against this and at the same time improves visualization, is to reconstruct the missing side by mirror reflection of the paired landmarks (Cardini, 2016, 2017). Thus, we estimated the right side by mirroring the left side paired landmarks and discarded the small asymmetries of the midplane landmarks, as described in the Supplementary Information of Cardini (2017). Overall, the correlations of size and shape data after estimating the missing side by mirror-reflection with the originals (midplane and left side landmarks only) were very high (respectively, 0.999 for centroid size, and 0.983 for the vectorized matrices of Procrustes shape distances). This means that, using only the left side landmarks, analytical results will be the same, but we still suggest using the data with the mirror-reconstructed right side for a better visualization.

Downloadable files included in GueSDat

For brevity, we henceforth refer to the left-side data with the missing side estimated by mirror-reflection and the midplane asymmetry removed simply as symmetrized data. Nonetheless, we stress that this is not the real symmetric component of a configuration with both left and right landmarks but is, in fact, just the left side information (plus the symmetrized midplane landmarks). Thus, symmetrized data are provided in four different formats:

A) The first is simple ASCII `txt` files: one file with the grouping variables and a second one with the the raw landmark coordinates in millimetres, both using the same identifier (Id) to match the specimens. The raw data allow users to compute size and shape variables using a Procrustes superimposition or, alternatively, for traditional morphometrics, interlandmark distances (i.e., linear measurements between pairs of landmarks) using, for instance, PAST (Hammer et al., 2001). As the original raw coordinates were recorded in millimetres, centroid size from Procrustes analysis and traditional interlandmark distances will all be in millimetres as well.

B) The second is the nts format, which is described in detail in the help file of the TPS Series (Rohlf, 2015). This is another type of `txt` file, with the extension changed to `nts`, and one of the oldest but still most commonly used formats in morphometrics. It only contains the raw landmark coordinates and the labels for the specimens and variables. As such, it provides less information than some other formats, but it is easy to manipulate, convert (e.g., in `csv` using `TPSUtil`;Rohlf, 2015) and can be directly imported into most GMM software, as well as into `R` (R Core Team, 2015). The labels for the specimens are the species code, followed by the sex code, followed by the identifier, with each number separated by an underscore. For instance, the first specimen is `1_0_426`, with `1` corresponding to *Allenopithecus nigroviridis*, `0` indicating that it is a female and `426` being the list identifier to retrieve more information from the file described in A) (e.g., the museum acronym and catalogue number for this specimen, which is USNM 395131). The labels for the variables are X, Y, Z followed by the landmark number (i.e., X1 Y1 Z1, X2 Y2 Z2).

C) The third format is a Morphologika file (O'Higgins and Jones, 2006). This is again a `txt` file and includes the identifier, the species classifier, the codes for sex and species, and also wireframes and polygons for the visualization. In the Morphologika format, a wireframe is a list of pairs of landmarks to be connected with a line (or link, in the terminology used by other GMM software) in the visualization of shape changes, and polygons are a list of triplets of landmarks used to draw surfaces, which can be used for rendering shape differences. As both wireframes and polygons are just visualization aids (Klingenberg, 2013b) and reflect an arbitrary choice, they can be modified by users as they deem appropriate. For the wireframe, this is very easily done using a graphical interface in either MorphoJ (Klingenberg, 2011) or `TPSUtil` (Rohlf, 2015).

D) Finally, the fourth format is the whole dataset including all classifiers and a sex covariate already imported into MorphoJ (Klingenberg, 2011). A MorphoJ project is an HTML file, which can be loaded in MorphoJ, one of the most widely used pieces of GMM software. MorphoJ is multi-platform, works using a graphical interface, is user-friendly and has a detailed manual. Besides a large number of analyses (from ordinations to regressions, analyses of covariation and modularity, comparative analyses and quantitative genetics, and so on), it also offers a number of simple tools for manipulating the data. For instance, using options in the "preliminaries" menu, one can include or exclude specimens or landmarks (both operations that can also be done using the `nts` file in `TPSUtil`), split the data into subsets according to a classifier and recombine them as appropriate, and average species and sexes. Data are also easily exported as `txt` files by selecting the appropriate branch of the project tree in which to save size and shape coordinates, principal components, results of regressions, and other outputs. The graphical output can be exported and modified for publications in other programs (e.g., using the `svg` format and the free multi-platform software Inkscape (https://inkscape.org/). In the MorphoJ project we also included the original left-side-only dataset (without mirror-reflected landmarks and symmetrized midplane) and one of the most recent molecular phylogenies for the group (downloaded in August 2016 from http://l0ktrees.nunn-lab.org/; Arnold et al., 2010).

Sample and geometric morphometric methods for current study

To address the scientific objectives of our current study, we used the data in GueSDat (see Tab. 1 and Fig. 1 for landmarks, and Tab. 2 for samples). For the comparison of variances between sexes, only species with at least five females and five males were included. Outlying specimens, clearly distant from most of the others (within sex and species) in box-plots and jitter plots of skull size and/or in phenograms and ordinations of shape data were excluded (see Viscosi and Cardini, 2011 and Cardini, 2013, for more detail on outlier detection).

Data were analysed with GMM using a Procrustes superimposition (Rohlf and Slice, 1990) to compute size and shape variables from the original raw Cartesian coordinates of the anatomical landmarks. Size was estimated using centroid size (the square root of the sum of squared deviations between each landmark and the baricenter of the complete landmark configuration). Skull centroid size is typically a good proxy for body mass in primates (Cardini et al., 2013, and references therein). Henceforth, for simplicity, we refer to centroid size simply as size, which was computed and the raw coordinates of each specimens di-

vided by the corresponding centroid size to remove size variation. Specimens were translated along the X, Y and Z axes so that their baricenters overlapped. This removes translational differences among individuals. Finally, rotational variation relative to the sample mean was minimized using a least square approach. This last operation completes the removal of irrelevant positional differences and produces a new set of coordinates, called Procrustes shape coordinates. The magnitude of shape differences in the resulting shape space can be quantified using the Procrustes shape distance, which is well approximated by the Euclidean distance between two shapes in the multivariate data space (or, more accurately, in the flat tangent space used to approximate the curved Procrustes shape space — with a very high accuracy in the GueSDat: $r_{\text{Procrustes distances - Euclidean distances}}$=0.999, computed in `TPSSmall`; Rohlf and Slice, 1990).

Statistical analyses

Before comparing variances between sexes, we tested if sexual dimorphism was significant. This was done using a 10000 permutation test for group mean differences, with the percentage of variance in size or shape explained by sex (R2) as the test statistic. This use of R2 is based on a regression of shape variables onto a dummy variable (i.e., a 0/1 code) for sex, and it is equivalent to using the Procrustes distance between the means of the two sexes. The observed R2 value was compared to the R2 distribution obtained by randomizing group (female versus male) affiliation within each species. If the observed R2 was larger than 95% of the R2 of the randomized samples, differences were considered significant at the 0.05 level. However, we also considered a more restrictive (generally over-conservative) Bonferroni-corrected threshold (0.0028), which takes into account that the same hypothesis was tested 18 times (i.e., once for each of the 18 larger species samples included in the analysis).

The equality of the magnitude in skull variance was tested using the permutational version of Levene's test, as explained in detail by Hallgrímsson et al. (2006); Cardini et al. (2007). Considering the current speculative nature of "Wainer's rule", and the possibility that male and female variance may be similar or even that variance is larger in females, we used a conservative two-tailed test, instead of simply testing if male variance was larger. As an estimate of the magnitude of multivariate shape variance we used the sum of the variances of the Procrustes shape coordinates. For variances, the range of uncertainties in estimates was also assessed by bootstrapping each sample (within species and sex) 1000 times, recomputing the variances and then computing the 2.5th and 97.5th percentiles of the bootstrapped distribution.

To explore the sensitivity of variances to sampling error in small samples, bootstrap estimates (1000 replicates) of variance in size and shape were also computed in subsamples of the two largest samples (*C. aethiops* and *C. mitis*). Thus, within each sex, the total sample was bootstrapped, subsamples were extracted at random using sample sizes in multiples of five (e.g., 5, 10, 15) and variances recomputed. Then, for each subsample (N=5, 10, 15 etc.), the mean variance and the 2.5 th and 97.5th percentiles of the distribution of bootstrapped variances, were computed and plotted against sample sample size.

Results and Discussion

Sexual dimorphism and comparison of variance between sexes

Results are shown in Tab. 2 and Fig. 2–4. The preliminary tests for sex differences in mean size and shape were, as expected in a group with well-documented sexual dimorphism (Cardini and Elton, 2008c, and references therein), highly significant. They remained significant even after an over-conservative Bonferroni correction, and indeed large proportions of variance were explained by sex in each species (from 50% to 90%, with an average of almost 80% for size, and from 8% to 45%, with an average of 21% for shape).

In the comparison of variance, there was generally good agreement between the results of the permutational version of the Levene's test and the overlap (or lack thereof) of the bootstrapped confidence intervals for

Table 2 – Sample composition (with species ordered according to increasing sample size), preliminary tests of mean shape differences (i.e., sexual dimorphism), and observed variances shown together with their ratios and the results of the Levene's tests. The ratio of female (F) to male (M) sample size (N) is shown with male-biased samples emphasized using a black background. Significant *p*s are in italics and those significant even if Bonferroni corrected for 18 tests ($p<0.0028$) are in underscored italics. Ratios of male to female variance are shown with species following the Wainer's rule emphasized using a black background; bold is used to show cases in which Wainer's rule is reversed and females have larger variance. A light grey background is instead used to show the few instances of bootstrap confidence intervals showing a disagreement with the results of the Levene's test.

	N				SIZE mean diff.		SHAPE mean diff.	SIZE variance					SHAPE variance								
Species	**F**	**M**	**tot.**	**F/M**	***p***	**R2**	***p***	**R2**	**F**	**C.I.**	**M**	**C.I.**	**M/F**	**Levene F**	***p***	**R2**	**F**	**C.I.**	**M**	**C.I.**	**M/F**	**Levene F**	***p***	**R2**
Allenopithecus nigroviridis	5	8	13	63%	*0.0010*	88.1%	*0.0006*	39.7%	87.0	(0–155.7)	287.0	(42.0–596.7)	3.3	1.774	0.2198	13.9%	0.00260	(0–0.002 603)	0.00178	(0.001 021–0.001 956)	**0.7**	5.610	*0.0302*	33.8%
Erythrocebus patas	9	18	27	50%	*0.0001*	89.8%	*0.0001*	45.2%	209.7	(32.4–326.1)	603.7	(259.1–944.1)	2.9	1.795	0.1945	6.7%	0.00247	(0.001 588–0.002 627)	0.00295	(0.002 351–0.003 119)	1.2	5.434	*0.0298*	17.9%
Cercopithecus hamlyni	15	15	30	100%	*0.0001*	89.2%	*0.0001*	26.8%	154.3	(58.7–252.0)	138.5	(51.3–226.6)	**0.9**	0.003	0.9626	0.0%	0.00260	(0.002 058–0.002 739)	0.00267	(0.002 025–0.002 926)	1.0	0.023	0.8807	0.1%
Miopithecus ogouensis	16	11	27	145%	*0.0001*	76.5%	*0.0001*	19.8%	64.0	(30.2–96.6)	44.2	(12.9–83.9)	**0.7**	1.615	0.2182	6.1%	0.00258	(0.002 028–0.002 795)	0.00259	(0.001 873–0.002 794)	1.0	0.049	0.8297	0.2%
Cercopithecus lhoesti	16	18	34	89%	*0.0001*	88.9%	*0.0001*	30.7%	178.5	(33.5–305.4)	204.2	(57.9–375.8)	1.1	0.006	0.9398	0.0%	0.00275	(0.002 162–0.002 982)	0.00239	(0.001 877–0.002 579)	**0.9**	2.048	0.1633	6.0%
Cercopithecus mona	16	19	35	84%	*0.0001*	88.9%	*0.0001*	25.2%	127.9	(45.9–227.6)	138.5	(78.4–193.1)	1.1	0.403	0.5222	1.2%	0.00287	(0.002 221–0.003 113)	0.00283	(0.002 388–0.002 907)	1.0	0.009	0.9238	0.0%
Cercopithecus petaurista	16	25	41	64%	*0.0001*	82.3%	*0.0001*	13.2%	55.1	(19.8–88.7)	101.2	(62.3–142.1)	1.8	2.476	0.1236	6.0%	0.00247	(0.001 970–0.002 618)	0.00280	(0.002 641–0.002 917)	1.1	3.873	0.0566	9.0%
Colobus guereza	17	16	33	106%	*0.0001*	66.3%	*0.0001*	9.7%	263.1	(134.5–366.7)	420.7	(185.3–652.7)	1.6	1.189	0.2901	3.7%	0.00268	(0.002 215–0.002 791)	0.00269	(0.002 105–0.002 869)	1.0	0.003	0.9546	0.0%
Cercocebus atys	23	21	44	110%	*0.0001*	86.3%	*0.0001*	25.1%	160.3	(82.8–228.0)	377.5	(167.0–576.8)	2.4	2.803	0.1001	6.3%	0.00239	(0.001 997–0.002 568)	0.00258	(0.002 151–0.002 737)	1.1	0.997	0.3253	2.3%
Cercopithecus neglectus	23	27	50	85%	*0.0001*	87.0%	*0.0001*	19.3%	115.1	(66.8–162.2)	190.2	(90.3–284.9)	1.7	0.872	0.3516	1.8%	0.00256	(0.001 938–0.002 680)	0.00257	(0.002 157–0.002 751)	1.0	0.001	0.9759	0.0%
Cercopithecus nictitans	24	23	47	104%	*0.0001*	49.9%	*0.0001*	11.3%	277.6	(104.6–442.0)	449.0	(245.1–616.4)	1.6	1.890	0.1750	4.0%	0.00299	(0.002 503–0.003 222)	0.00306	(0.002 921–0.003 180)	1.0	0.239	0.6260	0.5%
Cercopithecus cephus	29	29	58	100%	*0.0001*	73.5%	*0.0001*	7.8%	134.1	(73.3–197.1)	151.3	(86.9–212.2)	1.1	0.114	0.7382	0.2%	0.00323	(0.002 835–0.3391)	0.00295	(0.002 661–0.003 007)	**0.9**	2.675	0.1099	4.6%
Cercopithecus diana	31	32	63	97%	*0.0001*	81.9%	*0.0001*	13.6%	75.9	(38.3–112.0)	206.2	(107.2–314.5)	2.7	7.835	*0.0056*	11.4%	0.00293	(0.002 587–0.003 077)	0.00272	(0.002 377–0.002 975)	**0.9**	1.683	0.2051	2.7%
Cercopithecus campbelli	32	30	62	107%	*0.0001*	85.7%	*0.0001*	23.7%	74.3	(38.7–112.2)	230.7	(123.0–221.5)	3.1	8.056	*0.0065*	11.8%	0.00282	(0.002 477–0.002 959)	0.00275	(0.002 391–0.002 913)	1.0	0.198	0.6540	0.3%
Cercopithecus ascanius	37	39	76	95%	*0.0001*	75.9%	*0.0001*	12.7%	114.5	(64.0–166.7)	115.6	(65.3–169.9)	1.0	0.087	0.7668	0.1%	0.00301	(0.002 673–0.003 155)	0.00281	(0.002 508–0.002 974)	**0.9**	1.283	0.2635	1.7%
Cercopithecus pogonias	38	38	76	100%	*0.0001*	76.9%	*0.0001*	16.3%	126.3	(74.6–177.6)	141.1	(83.3–203.5)	1.1	0.172	0.6740	0.2%	0.00258	(0.002 331–0.002 711)	0.00264	(0.002 346–0.002 790)	1.0	0.132	0.7196	0.2%
Cercopithecus mitis	67	79	146	85%	*0.0001*	63.1%	*0.0001*	17.9%	364.8	(253.6–476.0)	776.7	(569.3–1000.0)	2.1	9.623	*0.0034*	6.3%	0.00284	(0.002 620–0.002 991)	0.00313	(0.002 907–0.003 286)	1.1	5.398	*0.0213*	3.6%
Chlorocebus aethiops	169	227	396	74%	*0.0001*	59.5%	*0.0001*	13.2%	280.2	(229.6–340.8)	585.6	(420.7–671.2)	2.1	20.119	*0.0001*	4.9%	0.00323	(0.003 050–0.003 331)	0.00322	(0.003 203–0.003 217)	1.0	0.066	0.7948	0.0%

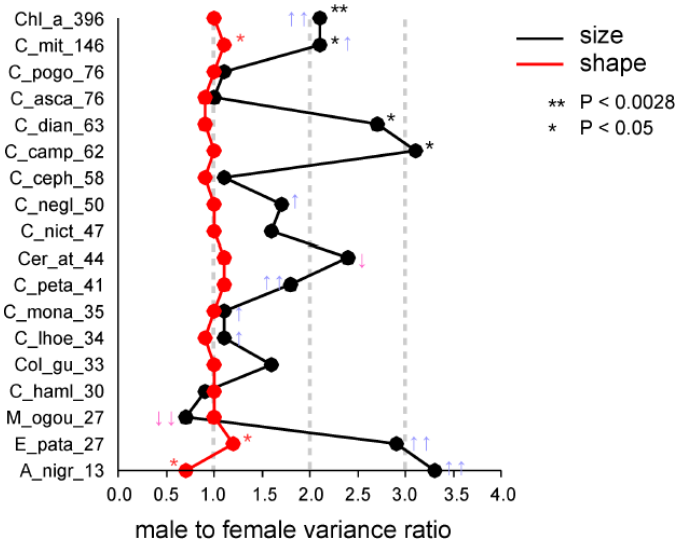

Figure 2 – Ratios of male to female size and shape variances. Species are ordered from the smallest to the largest sample (sample sizes appended to the abbreviated names). Abbreviations: A_nigr = *Allenopithecus nigroviridis*; E_pata = *Erythrocebus patas*; M_ogou = *Miopithecus ogouensis*; C_haml = *Cercopithecus hamlyni*; Col_gu = *Colobus guereza*; C_lhoe = *Cercopithecus lhoesti*; C_mona = *Cercopithecus mona*; C_peta = *Cercopithecus petaurista*; Cer_at = *Cercocebus atys*; C_nict = *Cercopithecus nictitans*; C_negl = *Cercopithecus neglectus*; C_ceph = *Cercopithecus cephus*; C_camp = *Cercopithecus campbelli*, C_dian = *Cercopithecus diana*; C_asca = *Cercopithecus ascanius*; C_pogo = *Cercopithecus pogonius*; C_mit = *Cercopithecus mitis*; Chl_a = *Chlorocebus aethiops*. Species with sample sizes where one sex exceeded the other by 10% were marked with one arrow; two arrows indicated species where sample size of one sex exceeded the other by ⩾20%. Upwarded pointing arrows = male-biased sample; downward pointing arrows = female-biased sample. Significant tests marked with asterisks.

the variance estimates. Most tests did not reach significance. In a few cases, in fact, the Levene's test was significant (but did not remain so after a Bonferroni correction), but the 97.5th percentile of the sex with the lower variance overlapped slightly with the 2.5th percentile of the sex with the higher variance. Even in these instances, however, the observed variance of that sex was always outside the confidence interval of the other sex (and vice versa). The only exception showing an apparently strong incongruence between the Levene's test results and the inference from the boostrapped confidence intervals was the very small sample of *Allenopithecus nigroviridis*: a consequence of having so few females (only five) was that some bootstrapped samples contained only replicas of the same individuals and, thus, showed no variance at all. In fact, with just five observations, the maximum number of different bootstrapped samples one can have is 756, and that means that almost 1/4 of *A. nigroviridis* bootstrapped samples appeared more than once in the computations, which are therefore inevitably unreliable. Also, in just one case, the comparison of shape variance in *Cercopithecus petaurista*, the Levene's test was not significant but close to the 0.05 threshold (p=0.0566), while the bootstrapped confidence intervals did not overlap (thus, suggesting appreciable differences). In summary, for both size and shape, most tests were not significant but, whereas male to female ratios of shape variances averaged 1, size variance ratios averaged almost 2.

Main trends

Overall, two main trends emerged from the tests of "Wainer's rule" in the guenons, *Cercocebus atys* and *Colobus guereza* skulls. The first was that despite rarely reaching significance, size variance was indeed larger in males than females most of the time. Only two species showed more variance in female size, and one species showed virtually identical variances in the two sexes. In all other cases (83% of the species), male variance was at least 10%, and up to more than three times, larger than female variance. In 60% of the 18 species used in the analysis, the difference was 50% or more, with almost 40% of the species having male variances at least twice as big as those in females. In some instances (e.g., *Miopithecus ogouensis*, with its very large female variance and many more females than males, and about half of the seven species with many more males than females and very large male variances [twice or more those of females]) sampling bias might have inflated estimates in one or the other sex. However, it seems unlikely that this alone explains the generally larger male variances: among the species with highest male to female variance ratios (>2), one of the strongly male-biased samples (*Chlorocebus aethiops*) had very large samples of both sexes (N>150), which should lead to fairly accurate estimates despite sampling bias (see below). More importantly, three of those species with very high male to female variance ratios had almost perfectly balanced samples or even, in one case (*C. atys*), a female-biased sample. Thus, although results need to be confirmed in future studies on larger and balanced samples, the consistency of the trend for size, and the fairly weak effect of the generally male-biased sampling, provide support for "Wainer's rule", at least in our sample.

The second main trend to emerge was that shape variance mostly showed very small differences between sexes. If any difference was found, there was no clear pattern and neither sex consistently showed larger variance. With the exception of the two smallest samples (*A. nigroviridis* and *Erythrocebus patas*, with total N<30), shape variances never differed more than 10% between the sexes, and only three samples reached significance (none if the significance threshold was Bonferroni-corrected). Indeed, half of the species had ratios of variance of about 1, and, of the other half, only four had variances slightly larger in males than in females, whereas in five it was the other way round, the opposite of "Wainer's rule". Thus, results strongly suggest a lack of important differences in the magnitude of shape variances between females and males. This is especially interesting if one considers that this happens regardless of the sex-bias in sample sizes, which, as with skull size, could have to some extent inflated estimates in the most represented sex. Even more surprising is that shape variances were similar despite often large differences in size variances and the generally pervasive effect of allometry: as shape tends to covary with size (as shown in skulls of a variety of mammals [e.g.,Cardini, 2017] as well as in guenons [Cardini and Elton, 2008c]), results of shape analyses often mirror those of size, unless allometry is controlled for (e.g., clinal variation in vervets, blue monkeys and red colobus [Cardini et al., 2007, 2010; Cardini and Elton, 2009]). Thus, we would have expected, in parallel with differences in size variances, larger differences in shape variances and a strong tendency to have more variance in males. In contrast, ratios of size and shape variances were uncorrelated (r=-0.007) and, in fact, more than half of the time they showed inconsistent trends (for instance, greater than 1 for size and less than or equal to 1 for shape or vice versa).

Sampling error

Figure 3 shows the results of the sensitivity of variance estimates to sampling, using the two largest taxonomic samples. Both samples were male-biased, as they included about 15–25% more males than females. This initial sampling bias could inflate male variance even in the perfectly balanced bootstrapped randomized subsamples. However, if there is a sample size threshold for which even the lower percentile (2.5th) of the estimates for the males and the upper percentile (97.5th) for the females do not overlap, that should provide strong evidence that differences in male and female variance are not simply due to sampling error.

Shape variances in male and female samples of *C. aethiops* were virtually identical with, in fact, males showing, in their total sample of 227 specimens, 0.3% less variance than found in 169 females. Thus, as males have the same variance as females despite a 25% larger sample, the issue of variance inflation becomes secondary: this is because either it does not exist or, if it did and male variance is overestimated, the real male variance would have been even lower and therefore clearly incongruent with "Wainer's ruleś'. When sample size is reduced using bootstrapped random subsamples, as expected the range of variance estimates increases, and it does so in a very similar fashion in both sexes (although the upper and lower percentiles of females are consistently slightly smaller than those of males). As long as sample sizes are larger than 100 for females and 150 for males, estimates of shape variances

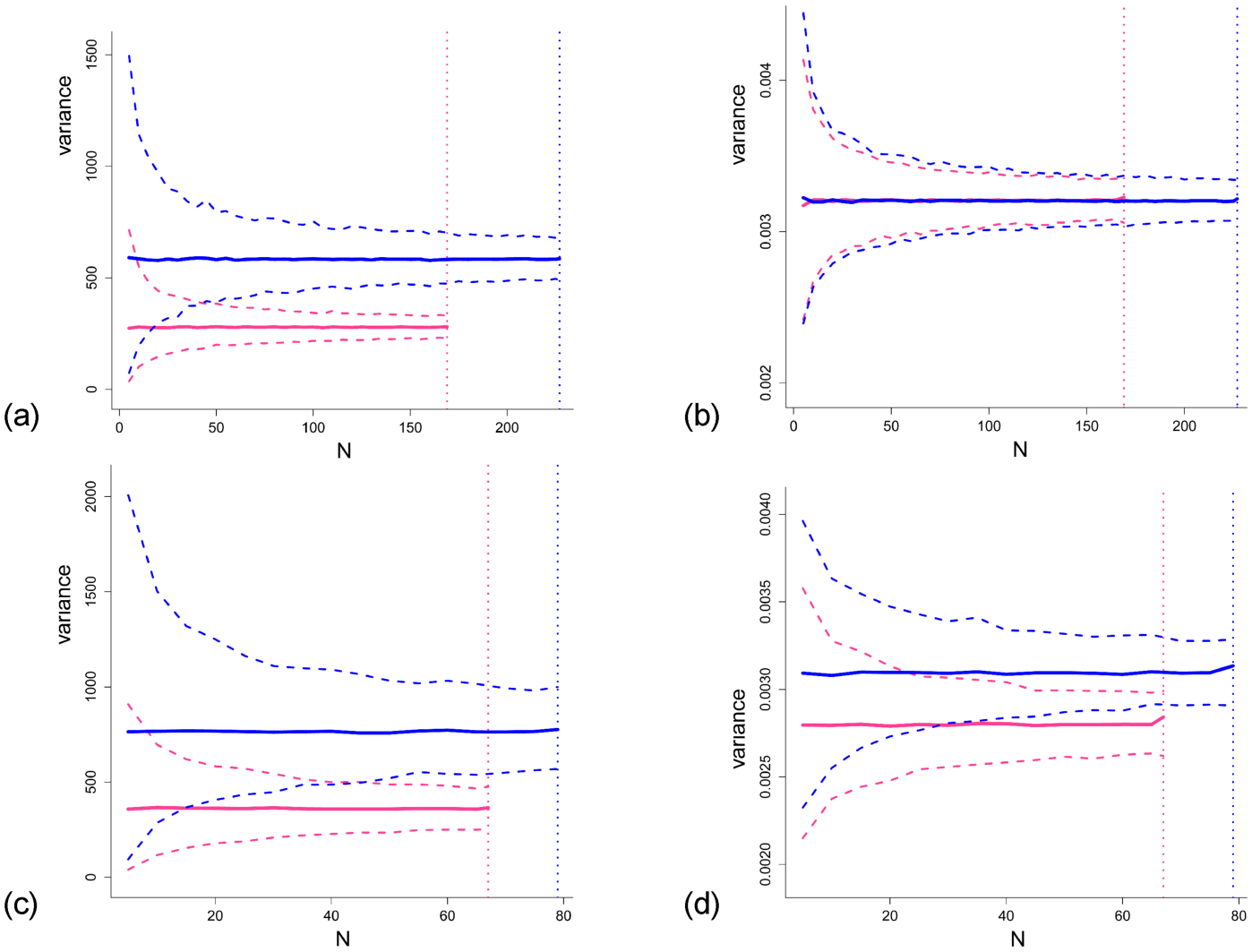

Figure 3 – Sensitivity of size (left side of the figure) and shape (right side) variance estimates to sampling error in females (pink) and males (blue). The mean (solid line), 2.5th and 97.5th percentiles (dashed lines) of 1000 bootstrapped random subsamples with sizes in multiples of five are shown, as well as the observed variances (total samples, marked by the dotted vertical lines) and its bootstrapped 2.5th and 97.5th percentiles: a–b) *C. aethiops*; c–d) *C. mitis*.

remain within about 5% of those observed in the total samples. With at least 30 females or 40 males, estimates range within 10% of the observed total sample values. Then, with less than about 20 individuals, uncertainties become rapidly larger (between approximately 15%, and up to 25% or more in samples of just 5 individuals). Thus, it seems that estimates of total shape variance are fairly reliable even when samples are not huge (ca. 30–40 individuals).

In *Cercopithecus mitis*, males have larger shape variance than females in the total sample. However, the total samples themselves are less than the minimum 100–150 specimens needed (based on the analysis of *C. aethiops*) to obtain estimates within 5% of the observed estimate. Thus, unsurprisingly, even the bootstrapped 2.5th–97.5th percentile range of the total sample is larger than 5% (being approximately 5–8% of the observed variance). Relatively low power, as well as the modest sex difference in variance in *C. mitis*, is probably why the 97.5th female upper percentile is above the male 2.5th lower percentile, which is in a sense congruent with the Levene's test being significant, but not highly so, and not significant at all if Bonferroni corrected.

In the bootstrapped random subsamples, uncertainties increase. No less than 30–35 individuals are needed for shape variances of female and male samples to vary up to, and no more than, about 10% of observed variance. This is strongly congruent with the results from *C. aethiops* that suggest a similar relative amount of uncertainty unless at least 30 individuals are available. With less than 25 specimens, shape variance estimates potentially become much more uncertain and can range up to about 25% with just five individuals per sex. This suggests that, if the pattern we found in the sensitivity analysis of *C. aethiops* and *C. mitis* is also valid in other species, overall findings from the seven species with about 30 or more specimens within each sex are reliable, at least with respect to the specific test and test statistics we are using. Also, of these seven species, five have almost perfectly balanced samples, and only the two largest samples (*C. aethiops* and *C. mitis*) are male-biased in terms of number of individuals. With only one exception, in all the 'reliable' samples, female and male variances are either similar or slightly larger in females. The exception is *C. mitis*, with significantly larger shape variance in males. However, this difference is small and, as mentioned, probably just marginally significant. Thus, it seems that the similarity of magnitude of skull shape variances in females and males is well supported in species of the GueSDat, with the exceptions due to sampling error (e.g., the two smallest species samples, *A. nigroviridis* and *E. patas*) or being relatively minor in extent (i.e., the +10% male variance of *C. mitis*).

In relative terms, the estimates of size variance are much more strongly affected by sampling than those of shape. Even in the total samples, for both *C. aethiops* and *C. mitis*, the upper (97.5th) and lower (2.5th) percentiles suggest potential inaccuracies of no less than ±15% (in fact, ca. ±25–30% in the smaller samples of *C. mitis*). This is approximately three times, or more, those observed for shape variance in the total samples. Further, as sample size is reduced in the resampling experiments, uncertainties in the estimates of size variance increase to become huge when only a few dozens of specimens are left. More precisely, in *C. aethiops*, the 2.5th–97.5th percentiles are about ±20% or less, compared to the total sample estimate, if at least 135 females or 160 males are available, but become about ±30% with 60–

65 specimens, and reach ±40% with 35 individuals in either sex, and ±120% when only five individuals are left. In *C. mitis*, when comparable sample sizes are available, results are virtually identical, with 2.5th–97.5th percentiles, compared to the observed total sample variances, being about ±30% with 65–70 specimens, ±40% with 35 individuals and about ±120% with just five specimens.

In both species, males have larger size variances but, when sample size of *C. aethiops* is 45 and *C. mitis* is 50, we start observing an overlap between the 2.5th male percentile and the 97.5th female percentile. This indicates that reliable results supporting a difference in size variances require very large samples (about 50 individuals or more) even when male variances are twice those of females. The sensitivity analysis also corresponds well with the results of the Levene's tests which were significant only in the largest samples (*C. aethiops* and *C. mitis*), or in moderately large ones (N>60) with very large ratios of variances (>2.5, as in *C. diana* and *C. campbelli*). Thus, the general trend showing larger male variances does support "Wainer's rule' but, until much larger samples are available, results should be seen as preliminary and must be interpreted with caution.

More generally, the results of the sensitivity analysis in this study correspond well to those by Cardini and Elton (2007) using the same *C. aethiops* sample but a simpler resampling procedure (i.e., random subsamples without bootstrapping). In that study too, in relative terms, differences in estimates of centroid size variance increased faster in smaller samples than differences in shape variance. In the female sample (Fig. 2b–c, Cardini and Elton, 2007), for instance, approximately the same 2.5th–97.5th percentiles of relative error in variance estimates (about ±40% of the variance observed in the total sample) were found in samples of 40 individuals for size but just 10 individuals for shape. Using teeth in Iceland horses, Cardini et al. (2015) also found a stronger effect of sampling error on estimates of variance in size: in samples of just five individuals, variances ranged from one-fifth to three times the value observed in the total sample for size, but were just ±50% of the total sample estimate for shape. Overall, therefore, all these studies suggest strongly that estimates of size variance should be taken particularly cautiously, while those of shape may be fairly precise even in samples of just a few dozens of individuals. It is important to bear in mind, however, that this conclusion refers only to the magnitude of shape variance estimated by the sum of the variances of the shape variables. Whether the same might be said for the full variance covariance matrix of shape data (which takes into account not only the total magnitude of shape variation but also the direction of shape changes) will require specific analyses. Nevertheless, our estimate that 30–35 individuals are required to obtain shape variances with a ±10% precision corresponds well with Polly (2005) suggestion, using landmark data on shrew teeth, that accurate estimates of variance covariance matrices require no less than 15–30 specimens, and with a similar suggestion (15–25 individuals) in Cardini et al.'s (2015) horse teeth study.

Implications and interpretations

Preliminary comment on homoscedasticity

If one sex does vary more than the other, an implication is that tests of sex differences assuming homoscedasticity inevitably violate this assumption. Based on our analyses of guenons, this issue seem to apply particularly strongly to size data, where males might indeed have more variance but estimates are also more strongly affected by sampling error and thus variance differences are often non-significant, likely because of low statistical power. Nevertheless, as there is a strong suspicion that the assumption of homogeneity of variance is violated in our own preliminary comparisons of female and male mean size, we re-ran (results not shown) all those tests using a t-test for samples with unequal variance, which confirmed (probably unsurprisingly given the magnitude of sexual dimorphism in this group) all the highly significant results we reported in Tab. 2.

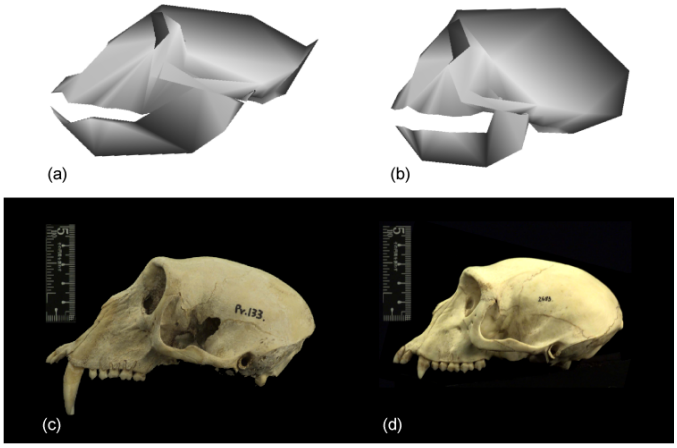

Figure 4 – Shape differences between male (a) and female (b) patas monkeys visualized in side view with a two-fold magnification using polygon surface rendering in Morphologika (O'Higgins and Jones, 2006). Pictures of male (c) and female (d) crania of *E. patas* (from http://lkai.dokkyomed.ac.jp/mammal/en/mammal.html – Takahashi et al., 2006).

Allometry and variance of specific anatomical regions

Exaggerated traits, such as ornaments or weapons under strong sexual selection (Bonduriansky, 2007a), might tend to show more variation than other traits, and that often occurs in parallel with large differences in body size. The large variation in these types of traits is often amplified by positive allometry. The highly variable horn of male horned beetles, for instance, shows a positive allometric relation to body size (i.e., it grows faster than body size; Johns et al., 2014), thus magnifying relative differences. Indeed, when there is a large variation in size, as in those beetles but also in our data on male guenons, the role of allometry on the proportions of different body parts should be carefully considered.

In this discussion, it is important to bear in mind the different analytical framework applied by Johns et al. (2014), as well as of those of many of the studies cited by Bonduriansky (2007a). The definition and framework used to assess allometry in those studies follow the Huxley-Jolicoeur's school, whereas our work, and GMM more generally, uses the Gould-Mosimann's approach Klingenberg (2016). This means that, when we talk about allometry in GMM, we refer broadly to the covariation of shape and size, without estimates of slope coefficients, as in the Huxley-Jolicoeur's framework. This is because GMM uses shape data, instead of contrasting the relative size of two traits, and also because, as with other types of coefficients, such as the loadings of principal components (Fig. 9, Viscosi and Cardini, 2011), coefficients of Procrustes shape coordinates cannot be meaningfully interpreted. Nevertheless, using Procrustes methods, one can still discuss whether a region of the total landmark configuration shows positive or negative allometry using the visualization. For instance, Fig. 4 exemplifies the typical pattern of sex differences in the skull of guenons, and many other primates, where the muzzle of the bigger males shows positive allometry (becoming relatively bigger), but the braincase, as well as the orbits, suggest negative allometry (being proportionally smaller than in females).

Despite the different approaches in studies of allometry, landmark data such as those contained in GueSDat, are flexible. One can extract specific linear measurements for bivariate tests, as those used in most of the analyses on insects, or subsets of landmarks measuring specific anatomical regions. As an example, we selected three–four landmarks measuring approximately the size of the muzzle (in ventral view, using landmarks 1, 10, 19 and 95), the size of the orbit (landmarks 45, 46, 47) and the braincase on the midplane (landmarks 19, 44, 63 and 64). If these landmarks capture the relative proportions of these regions with good approximation, further investigations of the effect of allometry on variance, as a function of the anatomical traits being measured, are possible. Thus, negative allometry in the orbit and braincase should produce smaller differences in size variance (i.e., make the male to female variance ratios smaller), while positive allometry should make them bigger in the muzzle. To a certain degree, this quick preliminary

exploration of regional size variances seems to support our predictions with, across species, average ratios of male to female variances of 1.2, 1.4 and 1.6 respectively in the orbit, midplane braincase and ventral muzzle.

We would probably have expected even smaller ratios for the braincase and larger ones for the muzzle. However, for simplicity in this example, we only employed very few of the landmarks available in the total configuration, and those that we selected may not have captured with complete accuracy the proportions of those regions. For instance, lengthening of the cranium is often accompanied by narrowing (i.e., the cranium becomes more dolichocephalic), an aspect which may be partially captured by the muzzle landmarks but not by the subset used for the braincase. In fact, when we regressed (not shown) the centroid sizes of these three regions against that of the total configuration (within each species, using pooled sexes), the averaged reduced major axis slopes suggested isometry for the muzzle (slope ≈1), moderate negative allometry for the braincase (slope ≈0.7) and a slightly stronger negative allometry for the orbit (slope ≈0.6). Crude as these exploratory analyses are, their results are congruent with the ratios of variances, which are on average about as big as with total cranial size using the muzzle landmarks, much smaller in the orbit, and intermediate in the braincase. At first glance, it may seem that the average isometric slope of the muzzle contradicts our previous statement that visualization indicates positive allometry in this region. However, because of differences in the landmarks employed (all 86 versus just 3–4) and, more importantly, in the framework (Huxley-Jolicoeur versus Gould-Mosimann) used to assess allometric variation, the results may not be directly comparable. Besides, as mentioned, the four muzzle landmarks simultaneously capture both lengthening and narrowing in bigger crania, thus affecting centroid size in opposite ways, with lengthening making it bigger and narrowing making it smaller.

Why do males have more variance in size but not in shape?

The results presented here, especially those concerning size, need to be interpreted with caution, because — as we have shown and discussed — very large random samples of females and males are required to obtain reliable estimates of variance. Unfortunately, these types of data are not easily obtained for mammals (especially large mammals) in museum collections. Nonetheless, there are strong hints that guenon total skull size may indeed be more variable in males than in females: the trend is consistent across virtually all species, and the fact that male-biased samples do not fully explain differences in variances suggests that the pattern may be real. If so, our findings fit well with the general notion (mostly supported by studies in invertebrates; Bonduriansky, 2007a,b; Bonneaud et al., 2016, and references therein) that secondary sexual characters may be more variable in the larger sex mainly because of plasticity in traits showing condition dependent responses. Although testing this prediction is well beyond the scope of our study, male guenons did not only vary more in skull size than females, but we also found (on average) more sex differences in variance in the muzzle, the most sexually dimorphic anatomical region among the three we analysed (muzzle averaged differences in size explained by sex ≈70% compared to 60% in the braincase and just 23% in the orbit).

If sex variance in guenon skull size seems to follow "Wainer's rule", and is congruent with predictions from studies in insects, shape consistently failed (with a few marginal exceptions) to show sex-related differences in the magnitude of variance. At first sight, this lack of differences looks particularly counter-intuitive, as at least some within- and between-sex shape variation in guenon skulls is allometric (Cardini and Elton, 2008c): one would therefore expect that a larger variance in size should be paralleled by a somewhat bigger shape variance. We suggest two, partly interrelated, potential reasons why this does not occur. The first is simply that allometry may indeed tend to have a pervasive effect on mammalian skulls, including the guenons (Cardini and Elton, 2008c; Cardini and Polly, 2013). However, as briefly anticipated in the previous paragraph, especially static allometry (i.e., intraspecific shape covariation with size within a precise ontogenetic stage, such as the adults) only actually explains a small amount of the total shape variance. For the data in GueSDat, within species and sex, this is on average about 9% (ranging from 5% to 21%; results not shown). Thus, even if allometry influences shape variance, its contribution may not be enough to produce an appreciable effect. The second reason, which in a sense is implicitly part of our first explanation, is the multivariate complex nature of shape, which may be less easy to change than size and thus less plastic and/or more resilient to evolutionary change, as briefly suggested in previous studies (Elton et al., 2010, as well as Cardini et al., 2013, and references therein; see Seetah et al., 2012 for an example in domestic species). Besides, size differences are probably much more crucial than shape variation in male-to-male competition, and this could increase the selective pressure to decouple size and shape in order for the former to respond rapidly in males to changes in condition (e.g., more and better food availability during growth) without impacting too much on shape. Condition dependence could thus be more evident in size than in shape, and more strongly affect the larger sex, which is always the male in Old World monkeys.

Conclusions

Our analysis suggests that plastic traits, such as size, and especially those under strong sexual selection, may indeed show more variation in males, when males are the bigger sex. This supports "Wainer's rule" for guenon skull size, and is in good agreement with studies on invertebrates. However, our detailed investigation of sampling error, although based on just two species, supports previous work that very large samples might be necessary in most instances to obtain accurate and robust findings, especially when size variance is estimated. For shape, sampling may be slightly less crucial, bearing in mind that we only addressed the issue of magnitude in variance, without exploring possible differences in the directions of variation and covariation. This notwithstanding, we hope that the promising findings obtained here might stimulate more research on "Wainer's Rule" and its applicability to other animals, across many phenotypic features.

Further, the analytical potential of the GueSDat is exemplified in this study, and this paper provides a reference that should be cited in publications originating from the use of GueSDat. The dataset is large both in terms of total sample size but also number of landmarks, and the taxonomic sampling within the guenons is extensive, as it includes almost all living species. The data are likely to be interesting not only for primatologists and mammalogists, but also for morphometricians interested in broader topics in evolutionary biology and in comparative studies, as well as for statisticians looking for real data on which to explore theoretical issues of statistical shape analysis.

References

Adams D., Rohlf F., Slice D., 2013. A field comes of age: geometric morphometrics in the 21st century. Hystrix 24(1): 7–14. doi:10.4404/hystrix-24.1-6283

Adler C.J., Donlon D., 2010. Sexual dimorphism in deciduous crown traits of a european derived australian sample. Forensic Sci. Int. 199: 29–37.

Arnold C., Matthews L.J., Nunn C.L. 2010. The 10kTrees website: A new online resource for primate phylogeny. Evol. Anthropol. 19: 114–118.

Bierry G., Le Minor J.-M., Schmittbuhl M., 2010. Oval in males and triangular in females? A quantitative evaluation of sexual dimorphism in the human obturator foramen. Am. J. Phys. Anthropol. 141: 626–631.

Bonduriansky R., 2007a. The evolution of condition-dependent sexual dimorphism. The American Naturalist 169: 9–19.

Bonduriansky R., 2007b. Sexual selection and allometry: a critical reappraisal of the evidence and ideas. Evolution 61: 838–849.

Bonneaud C., Marnocha E., Herrel A., Vanhooydonck B., Irschick D.J., Smith T.B, 2016. Developmental plasticity affects sexual size dimorphism in an anole lizard. Funct. Ecol. 30: 235–243.

Campbell N., Mahon R., 1974. A multivariate study of variation in two species of rock crab of the genus *Leptograpsus*. Aust. J. Zool. 22: 417–425.

Cardini A., 2013. Geometric morphometrics. Encyclopedia Of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO. Eolss Publishers, Paris, France. Available at www.eolss.net

Cardini A., 2016. Lost in the other half: improving accuracy in geometric morphometric analyses of one side of bilaterally symmetric structures. Syst. Biol. 65(6): 1096–1106. Data available at doi:10.560l:/dryad.mr2mh

Cardini A., 2017. Left, right or both? estimating and improving accuracy of one-side-only geometric morphometric analyses of cranial variation. J. Zool. Syst. Evol. Res. 55(1): 1–10. doi:10.1lll/jzs.12144 Data available at http://tinyurl.com/j8j3h6b

Cardini A., Dunn J., O'Higgins P., Elton S., 2013. Clines in africa: does size vary in the same way among widespread sub-saharan monkeys? J. Biogeogr. 40: 370–381.

Cardini A., Elton S., 2007. Sample size and sampling error in geometric morphometric studies of size and shape. Zoomorphol 126: 121–134.
Cardini A., Elton S., 2008a. Variation in guenon skulls (I): species divergence, ecdological and genetic differences. Journal of Human Evolution 54: 615–637.
Cardini A., Elton S., 2008b. Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. Biol. J. Linn. Soc. 93: 813–834.
Cardini A., Elton S., 2008c. Variation in guenon skulls (II): sexual dimorphism. Journal of Human Evolution. 54: 638–647.
Cardini A., Elton S., 2009. The radiation of red colobus monkeys (Primates, Colobinae): morphological evolution in a clade of endangered African primates. Zool. J. Linn. Soc. 157: 197–224.
Cardini A., Filho J.A.F.D., Polly P.D., Elton S., 2010. Biogeographic Analysis Using Geometric Morphometrics: Clines in Skull Size and Shape in a Widespread African Arboreal Monkey. In: Elewa A.M.T. (Ed.) Morphometrics for Nonmorphometricians. Lecture Notes in Earth Sciences, vol. 124. Springer, Berlin, Heidelberg.
Cardini A., Polly P.D., 2013. Larger mammals have longer faces because of size-related constraints on skull form. Nature Commun. 4: 2458. doi:10.1038/ncomms3458
Cardini A., Seetah K., Barker G., 2015. How many specimens do I need? Sampling error in geometric morphometrics: testing the sensitivity of means and variances in simple randomized selection experiments. Zoomorphol. 134(2): 149–163.
Cardini A., Thorington R.W. Jr., Polly P.D., 2007. Evolutionary acceleration in the most endangered mammal of Canada: speciation and divergence in the Vancouver Island marmot (Rodentia, Sciuridae). J. Evol. Biol. 20: 1833–1846.
Dunbar R.I.M., 1990. Environmental determinants of intraspecific variation in body weight in baboons (*Papio* spp.). J. Zool. 220: 157–169.
Eickwort K.R., 1969. Differential variation of males and females in *Polistes exclamans*. Evolution 3: 391–405.
Elton S., Dunn J., Cardini A., 2010. Size variation facilitates population divergence but does not explain it all: an example study from a widespread african monkey. Biol. J. Linn. Soc. 101(4): 823–843.
Forsman A., 1991. Variation in sexual size dimorphism and maximum body size among adder populations: effects of prey size. J. Anim. Ecol. 60: 253–267.
Gaston K.J., Chown S.L., Evans K.L., 2008. Ecogeographical rules: elements of a synthesis. Journal of Biogeography. 35: 483–500.
Gonzalez P.N., Bernal V., Perez S.I., 2011. Analysis of sexual dimorphism of craniofacial traits using geometric morphometric techniques. Int. J. Osteoarchaeol. 21: 82–91.
Grubb P., Butynski T.M., Oates J.F., Bearder S.K., Disotell T.R., Groves C.P., Struhsaker T.T., 2003. Assessment of the Diversity of African Primates. Int. J. Primat. 24: 1301–1357.
Hallgrímsson B., Brown J.J.Y., Ford-Hutchinson A.F., Sheets H.D., Zelditch M.L., Jirik F.R., 2006. The brachymorph mouse and the developmental-genetic basis for canalization and morphological integration. Evolution and Development 8: 61–73.
Hammer O., Harper D., Ryan P., 2001. PAST: Paleontological statistics software package for education and data analysis. Paleontol. Electron. 4(1): 1–9.
Johns A., Gotoh H., McCullough E.L., Emlen D.J., Lavine L.C., 2014. Heightened condition-dependent growth of sexually selected weapons in the Rhinoceros Beetle, *Trypoxylus dichotomus* (Coleoptera: Scarabaeidae). Integr. Comp. Biol. 41: 614–621.
Klingenberg C.P., 2011. MorphoJ: an integrated software package for geometric morphometrics. Mol. Ecol. Resour. 11: 353–357.
Klingenberg C.P., 2013a. Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. Hystrix 24(1): 15–24. doi:10.4404/hystrix-24.1-7691
Klingenberg C.P., 2013b. Cranial integration and modularity: insights into evolution and development from morphometric data. Hystrix 24(1): 43–58. doi:10.4404/hystrix-24.1-6367
Klingenberg C.P., 2016. Size, shape, and form: concepts of allometry in geometric morphometrics. Dev. Genes Evol. 226: 113–137.
Klingenberg C.P., Barluenga M., Meyer A., 2002. Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. Evolution 56: 1909–1920.
Leutenegger W., Masterson T.J., 1989. The ontogeny of sexual dimorphism in the cranium of Bornean rang-utans (*Pongo pygmaeus pygmaeus*): I. Univariate analyses. Zeitschrift für Morphologie und Anthropologie 78: 1–14.
Marcus L.F., 1990. Traditional morphometrics. Proceedings of the Michigan morphometrics workshop 2: 77–122.
Moczek A.P., 2006. Pupal remodeling and the development and evolution of sexual dimorphism in horned beetles. Am. Nat. 168: 711–729.
O'Higgins P., Jones N., 2006. Morphologika 2.2. Tools for shape analysis. Hull York Medical School, University of York, York. Avalaible at https://sites.google.com/site/hymsfme/resource
Polly P.D., 2005. Development and phenotypic correlations: the evolution of tooth shape in *Sorex araneus*. Evol. Devel. 7: 29–41.
Power C., Lake J.K., Cole T.J., 1997. Body mass index and height from childhood to adulthood in the 1958 British born cohort. The American Journal of Clinical Nutrition 66: 1094–1101.
Przybylo R., 1995. Intersexual niche differentiation: field data on the great tit *Parus major*. J. Avian Biol. 26: 20–24.
Quinn T.P., Foote C.J., 1994. The effects of body size and sexual dimorphism on the reproductive behaviour of sockeye salmon, *Oncorhynchus nerka*. Anim. Behav. 48: 751–761.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/
Rohlf F.J., 2015. The tps series of software. Hystrix 26(1): 9–12. doi:10.4404/hystrix-26.1-11264
Rohlf F.J., Slice D., 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. Syst. Biol. 39: 40–59.
Seetah T.K., Cardini A., Miracle P.T., 2012. Can morphospace shed light on cave bear spatial-temporal variation? Population dynamics of *Ursus spelaeus* from Romualdova pećina and Vindija, (Croatia). J. Archaeol. Sci. 39: 500–510.
Stigler S.M, 1980. Stigler's law of Eponymy. Transactions of the New York Academy of Sciences 39: 147–157.
Tague R.G., 1989. Variation in pelvic size between males and females. Am. J. Phys. Anthropol. 80: 59–71.
Takahashi H., Yamashita M., Shigehara N., 2006. Mammalian Crania Photographic Archive, Second Edition (MCPA2). Revised on December 1, 2005. Available at http://lkai.dokkyomed.ac.jp/mammal/en/mammal.html
Teder T., Tammaru T., 2005. Sexual size dimorphism within species increases with body size in insects. Oikos 108: 321–334.
Viscosi V., Cardini A., 2011. Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. PLoS ONE 6: e25630. doi:10.1371/journal.pone.0025630
Wainer H., 2007. The most dangerous equation. Am. Sci. 95: 249–256.

Associate Editor: P. Colangelo

Supplemental information

Additional Supplemental Information may be found in the online version of this article:

Supplemental data S1 ASCII files with the grouping variables and the raw landmark coordinates.
Supplemental data S2 Landmark coordinates in `nts` format.
Supplemental data S3 Morphologika file.
Supplemental data S4 Whole dataset in MorphoJ format.