

Sample size and sampling error in geometric morphometric studies of size and shape

Andrea Cardini · Sarah Elton

Received: 24 October 2006 / Accepted: 8 June 2007 / Published online: 17 July 2007
© Springer-Verlag 2007

Abstract Geometric morphometric studies are increasingly becoming common in systematics and palaeontology. The samples in such studies are often small, due to the paucity of material available for analysis. However, very few studies have tried to assess the impact of sampling error on analytical results. Here, this issue is addressed empirically using repeated randomized selection experiments to build progressively smaller samples from an original dataset of ~400 vervet monkey (*Cercopithecus aethiops*) skulls. Size and shape parameters (including mean size and shape, size and shape variances, angles of allometric trajectories) that are commonly used in geometric morphometric studies, are estimated first in the original sample and then in the random subsamples. Estimates are then compared to give an indication of what is the minimum desirable sample size for each parameter. Mean size, standard deviation of size and variance of shape are found to be fairly accurate even in relatively small samples. In contrast, mean shapes and angles between static allometric trajectories are strongly affected by sampling error. If confirmed in other groups, our findings may have substantial implications for studies of morphological variation in present and fossil species. By performing rarefaction analyses like those presented in our study, morphometricians can be easily provided with important clues

on how a simple but crucial factor like sample size can alter results of their studies.

Keywords Accuracy · Geometric morphometrics · Primates · Randomization · Sampling error

Introduction

When working with relatively small skeletal and palaeontological samples, the issue of adequate sample size is often troublesome. Sample sizes for many morphometric studies of living and fossil mammals are determined on the basis of the number of specimens available for study rather than on an a priori judgement of what is optimal. In vertebrate palaeontology, it is almost a truism that the limited number of specimens available for analysis can restrict the inferences that can be made about palaeobiology and evolutionary history. Nonetheless, very few would argue that analysis of skeletal or palaeontological samples should not be undertaken simply because of the relatively small numbers of specimens that can be obtained. It is therefore important to understand when sample size matters, and to be able to appreciate how sample size might influence the results of palaeontological and morphometric studies.

Different types of studies, and indeed different types of data, require distinct sampling strategies. Descriptions of recently discovered fossil specimens and the naming of new taxa are often based on a small number of individuals. Having access to a comparative sample is essential, but such samples tend to be relatively small, and the conclusions drawn about the general population are usually restricted to defining which features are present or absent in a particular taxon. In such cases, the sampling strategy might not be critical to the outcomes of the work. However, much

A. Cardini (✉)
Museo di Paleobiologia e dell'Orto Botanico,
Università di Modena e Reggio Emilia,
via Università 4, 41100 Modena, Italy
e-mail: alcardini@interfree.it; cardini@unimo.it

S. Elton
Hull York Medical School, The University of Hull,
Cottingham Road, Hull HU6 7RX, UK
e-mail: sarah.elton@hyms.ac.uk

of the research conducted in palaeontology, and virtually all morphometric studies on present species, goes beyond the qualitative description of anatomical characters to infer population parameters. When selecting a sample for inferential purposes, a judgement must be made as to whether the data sampled reflect the variability within the population as a whole, and if it is possible to make generalizations on the basis of the chosen sample. The reliability and robusticity of the conclusions will depend on the type of statistical descriptor or test used, and the shape of the data. Studies of many vertebrate species must take sexual dimorphism into consideration, for example, and palaeontological data are additionally complicated by the effects of taphonomy and often highly uneven distributions in time and space. Consideration of how sample size might influence results and conclusions is thus highly desirable for much, if not all, of the inferential and predictive work that is conducted in palaeontology but also for quantitative morphological comparisons of modern species.

With the exception of rarefaction analyses for assessing variation in measures of disparity within a clade as a function of the number of available species, the extent to which analytical results are affected by sample size in morphometric data has been examined in relatively few studies. Studies on how sampling error might affect estimates of statistical parameters are virtually absent in geometric morphometrics, an innovative, statistically powerful and visually effective method for quantitative comparisons of forms, which has emerged as the leading branch of morphometrics in the last two decades (Rohlf and Marcus 1993; Adams et al. 2004; Slice 2005). The few studies that have been done have often used simulations and mathematical modeling to examine the issues theoretically. For instance, Rohlf (2000a, 2003a) compared statistical power and accuracy of mean shape estimates produced by different geometric morphometric methods as a function of sample size using configurations of three landmarks built by adding isotropic errors around coordinates of either an isosceles or an oblique triangle. The simulation showed that Procrustes methods generally perform better than other methods but errors in samples consisting less than 100 specimens can be two to six times larger than in samples consisting of 500 specimens (Fig. 6, p. 676). Empirical approaches, using real morphological data, are relatively rare, but those that have been conducted indicate that sample sizes significantly influence results. In the research on ontogenetic shape trajectories in African apes, repeated randomized selection experiments indicated that small sample sizes (<15–20 individuals of mixed ages) provided unreliable estimates of trajectories, with consequent implications for estimating trajectories in fossil hominins (Cobb and O'Higgins 2004). Random subsamples were used also by Polly (2005) in a study on shrew teeth mor-

phology to examine the effect of small sample size on comparisons of variance–covariance matrices. It was found that matrix structure varied significantly in samples of less than 15–30 specimens. Both these studies demonstrate that even though morphometricians are often forced to work with relatively small samples, they should be aware of how their sampling might impact on the conclusions that can be drawn.

The aim of the study reported here is to test empirically rather than theoretically how altering sample size influences estimates of population parameters that are commonly used in geometric morphometric studies, using data from modern cercopithecine monkeys. A large sample ($N \approx 400$) of adult *Cercopithecus aethiops* (Linnaeus, 1758) skulls was measured using three-dimensional Cartesian coordinates of anatomical landmarks. Parameters (mean and standard deviation of size, mean shape and shape variance, angles of static allometric trajectories, goodness of fit of different regression models) were computed in the original sample, in its bootstrap replicas and in random subsamples with progressively smaller sample size ($N = 100, 90, 80$ and so on). The effect of sample size on parameters is examined by comparing parameters in subsamples with those of the original sample. Trends in variation of parameters in relation to sample size are investigated and compared to provide clues about potential issues with inadequate sampling in geometric morphometric studies of skull shape variation in cercopithecine monkeys. This approach allows to directly assess the effect of sampling error on results from a morphometric study of a real sample, and represents a kind of exploratory analysis which is easily applicable whenever a large sample is available for at least one of the study taxa. Potential implications for taxonomic and palaeontological analyses of form, which are often concerned with small variation among closely related taxa for which large samples are seldom available, will be discussed.

Materials and methods

Data collection

The sample comprised 396 adult specimens of *C. aethiops* (Linnaeus, 1758), of which 169 were females and 227 were males. 146 specimens of *C. mitis* Wolf, 1822 (67 females and 79 males) were also included in the analysis of static allometry. The maturity of each specimen was judged on the basis of full eruption of third molars and canines. Specimens came from the collections of the National Museum of Natural History (Washington, DC, USA), American Museum of Natural History (New York, NY, USA), Museum of Comparative Zoology of the University

of Harvard (Cambridge, MA, USA), Field Museum of Natural History (Chicago, IL, USA), Museum für Naturkunde of the Humboldt University (Berlin, Germany), Zoologische Sammlung des Bayerischen Staates (Munich, Germany), Royal Museum for Central Africa (Tervuren, Belgium), British Museum of Natural History (London, UK), Powell-Cotton Museum (Birchington, UK).

Three-dimensional coordinates of anatomical landmarks were directly collected by the same person on crania and mandibles using a 3D-digitizer (MicroScribe 3DX, Immersion Corporation, San Jose, Ca, USA). Landmarks were digitized only on the left side to avoid redundant information in symmetric structures. The set (configuration) of 86 landmarks used for the analysis is shown in Fig. 1a. Definitions of landmarks can be found in Cardini et al. (2007b).

Landmarks on crania and mandibles were digitized separately. Three registration points were digitized on pieces of plasticine stuck on the two condyles and below the incisors of the mandible of each specimen. These landmarks were recorded twice: first, on the mandible articulated to the cranium (after the digitization of the cranial landmarks) and, then, on the disarticulated mandi-

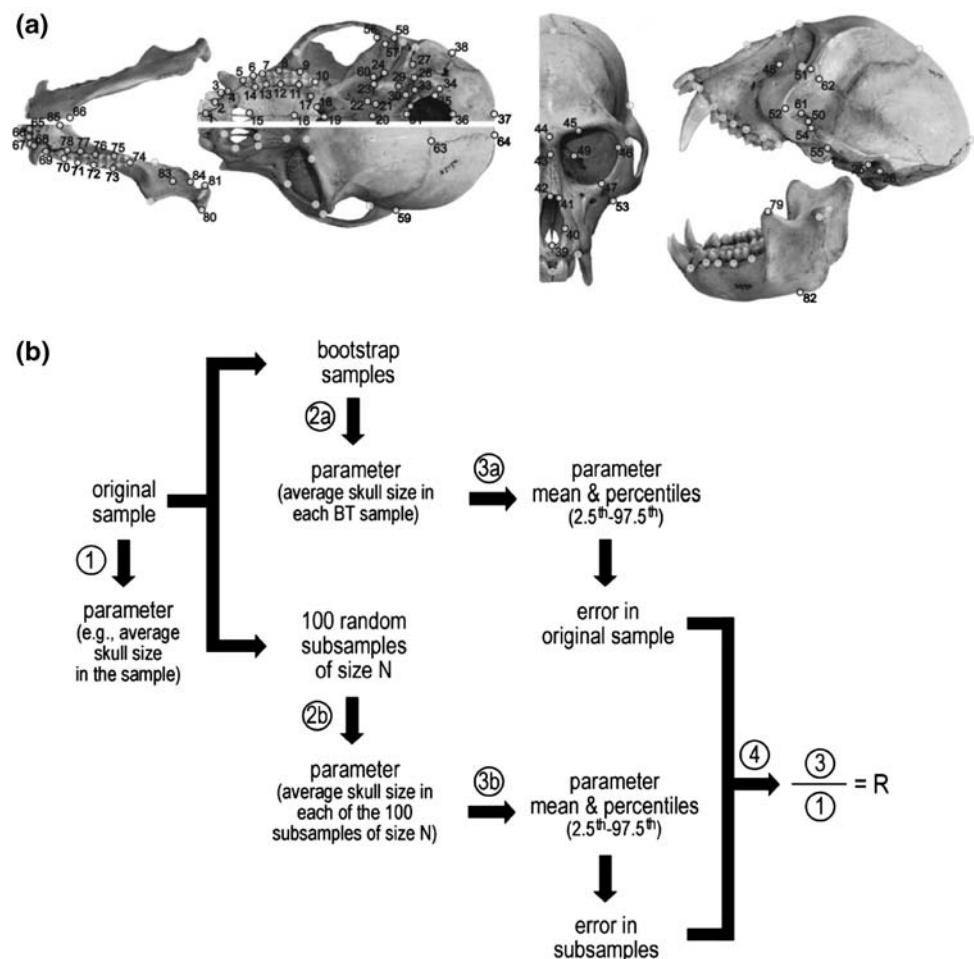
ble (after the digitization of the mandibular landmarks). The three registration points were chosen in the form of a large triangle with distant vertices in order to minimize the measurement error relative to the size of the triangle. Data collected on the mandible were then aligned onto the same coordinate system as those collected on the cranium by applying a least-squares superimposition (see below) of the three points so that the rigid rotation derived from them applies to all landmark coordinates. Software written in Visual Basic (N. Jones, unpublished data) was used for this. The three landmarks used for matching the cranium and mandible configurations were eventually discarded and only the 86 anatomical landmarks are used in the analyses.

Measurement error and estimates of a small number of missing landmarks (1–2 in 13.7% and 3–6 in 3.1% of specimens) were described in A. Cardini and S. Elton (submitted data) and shown to have negligible effects on the analysis.

Geometric morphometrics

Analyses were performed using geometric morphometrics (Rohlf and Marcus 1993; Adams et al. 2004) in the

Fig. 1 **a** Landmark configuration. **b** Diagram summarizing steps of the analysis for computing and comparing parameters in samples of different sizes (R ratio between subsample (or bootstrap) sample parameter and original parameter)



following computer programs: Morpheus (Slice 1999), Morphologika (O'Higgins and Jones 2006), TPSSmall 1.20 (Rohlf 2003b), NTSYS-pc 2.2d (Rohlf 2005). The form of an organism (or its organs) is captured by the Cartesian coordinates of a three-dimensional configuration of anatomical landmarks. Differences in landmark coordinates due to the position of the specimens during the digitization process are removed, and size is standardized. This was achieved in the present study by optimally superimposing landmark configurations using a process called generalized Procrustes analysis (GPA), which is based on a least-squares algorithm (Rohlf and Slice 1990). Centroid size (henceforth, simply called 'size', for brevity) is a measure of the dispersion of landmarks around their centroid and it is computed as the square root of the sum of squared distances of all landmarks from the centroid. The new Cartesian coordinates obtained after the superimposition are the shape coordinates used for statistical comparisons of individuals. The shape differences between landmark configurations of two individuals can be summarized by their Procrustes distance (PRD), which is approximately the square root of the sum of squared distances between pairs of corresponding landmarks. Henceforth, we will refer to Procrustes shape distances by simply using the term shape distances or the abbreviation PRD.

Variations in the form of the landmark configurations were examined using Procrustes-based geometric morphometrics, rather than angle- or distance-based approaches, because geometric morphometrics more precisely estimate the mean, have lower type I error rates, higher statistical power and lower bias in simulations based on the assumption of independent isotropic distributions of landmarks (Rohlf 2000a, b, 2003a). Although this assumption is unrealistic, there are good theoretical reasons (Dryden and Mardia 1993; O'Higgins 1999) and numerous published studies that suggest deviations from isotropic distributions, which still lead to reasonable mean estimates and preserve statistical power as long as shape variations are small with respect to all possible configurations of landmarks. This is a common situation in biological data and very much the case with the present data. Thus, while Procrustes methods can never faithfully represent landmark 'movements' they present a robust means of estimating form variants within a sample, ordinating these with respect to each other in a way that well reflects the underlying biology and testing for significant differences in the form of landmark configurations taken as a whole. Further by applying appropriate mathematical functions (e.g., the thin plate spline) to warping of images or grids they allow localization of shape differences between pairs of landmark configurations.

An extensive introduction to applications of geometric morphometrics in biology is provided by Zelditch et al.

(2004). Detailed mathematical descriptions of geometric morphometric methods are available in Bookstein (1991) and Dryden and Mardia (1998). Guidelines on how to implement linear statistical models in geometric morphometrics can be found in Rohlf (1998) and Klingenberg and Monteiro (2005).

Statistics

The parameters computed for the original sample (including all individuals, split by sex as appropriate), bootstrap samples or random subsamples were:

- (1) mean size;
- (2) standard deviation of size;
- (3) mean shape (i.e., the mean of all specimens in a sample after the superimposition);
- (4) total shape variance (computed as the sum of variances of all shape variables);
- (5) angle between allometric trajectories (see below) of female and male *C. aethiops* (within species static allometries) or of *C. aethiops* and *C. mitis* with pooled sexes (between species static allometries);
- (6) percentage of shape variance explained by size (see below).

Statistical analyses were performed using NTSYS-pc 2.2d (Rohlf 2005) and SPSS 11.5.0 (2004). Each parameter was computed first in the original sample of *C. aethiops* (Fig. 1b₁), split by sex due to the large degree of sexual dimorphism observed in *C. aethiops* and common to the guenons as a whole (A. Cardini and S. Elton, submitted data). Females and males were pooled only in the comparison of static allometries of *C. aethiops* and *C. mitis*.

In the second step of the analysis, the variation in parameters of the original sample was estimated using bootstraps (Fig. 1b_{2a}). Thus, 100 (including observed) samples of the same size as that of the original were built by resampling with replacement the original sample. Parameters were computed for each bootstrap sample and the average and 95% confidence interval (2.5th and 97.5th percentiles) of each parameter were calculated. Then (Fig. 1b_{2b}), 100 subsamples of a given size were built by random selection of specimens from the original sample. For instance, for $N = 10$, ten individuals were randomly drawn from the original sample and this was repeated 100 times. Thus, 100 subsamples of ten individuals were built and parameters were computed for each subsample. As for bootstraps, the average and the 2.5th and 97.5th percentiles of each parameter were calculated. This procedure was repeated in subsamples of $N = 10, 20, 30, \dots, 80, 90, 100$.

All analyses of allometric variation were done using multivariate regressions of all shape variables onto size. The natural logarithm of size was used, following Mitter-

oecker et al. (2005). Thus, vectors (trajectories) of shape variation predicted by size were computed within each group and angles between pairs of vectors (e.g., females and males) were calculated as the dot product of regression coefficients. Small angles imply (almost) parallel trajectories and large angles imply divergent trajectories.

The percentage of shape variance explained by size was computed as:

$$100 \times [1 - (\text{shape variance of the regression residuals} / \text{total shape variance})].$$

It is an overall measure of fit of the regression, that can be used for comparing results of fitting different sets of independent variables (size, ecological variables, geographic coordinates and so on) or results of regressions of the same variables in different samples (e.g., percentages of allometric shape in different species). This measure can also be used to compare the goodness of fit of different regression models. For instance, clues about the amount of divergence between two trajectories can be provided not only by angles but also by comparing percentages of explained variance obtained by fitting separate lines (independent slopes), parallel lines (separate lines with identical slopes) or the same line through both samples. If different models fit the samples equally well, it implies either that trajectories are not divergent (parallel or overlapping) or that the variation in the direction set by the independent variable is not much bigger than that in any other direction, and those trajectories are thus not particularly meaningful. The statistical significance of the regression of shape onto size was first tested in each original sample (and found to be highly significant—results not shown), and then trajectories of either female and male *C. aethiops* or those of *C. aethiops* and *C. mitis* were compared using the three regression models described above. For each model, the effect of sampling error on the percentages of explained variance was examined.

To make results comparable and easier to interpret, the ratio $R = P_N/P_{\text{obs}}$ was computed and plotted against sample size. P_N is the parameter in a subsample of a given size, and P_{obs} is the parameter in the original sample. For instance, if the observed angle between two trajectories is 15° and the same angle in a subsample of 10 specimens is 45° , R is $45^\circ/15^\circ = 3$, which means that the angle in the subsample is three times larger than in the original sample. The average of R and the 2.5th and 97.5th percentiles of R were computed, and their variation as a function of sample size was illustrated with profile plots using black lines for the average R and grey lines for the percentiles. Thus, a profile plot with a horizontal black line suggests accuracy (P_N close to P_{obs}), while

close and almost parallel grey lines indicate precision (P_N close to the average of a given sample size). On the other hand, a pronounced curvature in the black line and a progressive increase in the distance of the grey lines indicate low accuracy and low precision.

To measure how close subsample mean shapes were to the best available estimate of the mean shape, all means were superimposed and their PRD to the observed mean computed, as described in Rohlf (2003a, b). The average and the 95th percentile of these distances (D_N) across subsamples of a given size (e.g., $N = 10$) were used to summarize how similar the observed mean was to the means in a sample of a particular size. To make results comparable and aid interpretation, $R_D = D_N/D_{\text{obs}}$ was computed and plotted against sample size. Profile plots were drawn following the same conventions as for R . In this case, D_{obs} was either the average PRD between bootstrap mean shapes and the observed mean (a measure of the uncertainty around the observed mean shape) or the PRD between the observed mean shape of *C. aethiops* and that of *C. mitis* (a measure of interspecific variation between well separated and ecologically divergent guenons). For instance, the average PRD of female means to the original mean shape was 0.01810 for subsamples of ten females, 0.00438 for bootstrap samples and 0.04940 from the mean of *C. aethiops* to the mean of *C. mitis*. Thus, R_D is $0.01810/0.00438 = 4.1$, relative to the uncertainty in the observed mean shape, and it is $0.01810/0.04940 = 0.4$, relative to the interspecific divergence. In the first case (within species), it implies that the distance of mean shapes of samples of ten females to the observed mean was on average about four times larger than the uncertainty in the estimate of the observed mean. In the second case (interspecific), it shows that the distance of mean shapes of samples of ten females to the observed mean was on average as large as 40% of the interspecific shape distance between *C. aethiops* and *C. mitis*.

The procedure for computing and comparing parameters in different-sized samples is summarized in Fig. 1b. Assuming that estimates of parameters in the original samples are reasonably close to the true parameters in the population, a parameter is accurately estimated in subsamples of size N when their estimates are on average close to the observed. Percentiles are used to provide information on how close estimates of parameters in subsamples of a given size are on average to the observed. They are also used to show how variable the distance between observed and subsample estimates is. Parameter estimates are precise if their distances to the observed show little variation (i.e., they are close to the average). However, this use of the term precision does not strictly correspond the conventional definition of statistical precision, and is only adopted to make a semantic distinction between how accurate

subsample estimates are on average and how their accuracy varies among subsamples.

Results

Results for females and males are very similar in all split-sex analyses. Thus, profile plots of parameter ratios (R or R_D) are shown for females only. The tabulated results are summarized using parameter means and percentiles.

The accuracy and precision of estimates of size parameters, including mean and standard deviation of size computed in the original sample, bootstrap samples and subsamples, are shown in Table 1 and Fig. 2a–b. Estimates of the average size of *C. aethiops* are both accurate and precise. Even in samples of ten individuals, 95% of estimates are within a range of 0.95–1.05 times the observed. The standard deviation of size also shows modest differences compared to the observed value. However, with less than 30 individuals, the range of standard deviations in subsamples becomes much wider and the standard deviation

can be up to about 0.5 smaller or 1.3 larger than the observed. Shape variance in a sample also tends to be accurate and precise (Table 2, Fig. 2c). Even in the smallest subsamples, 95% of shape variances are within a range of 0.75–1.25 times the observed. In contrast, mean shape variation around the observed mean can be fairly large (Table 2, Fig. 2d–e). Subsample mean shapes rapidly become more distant from the observed mean when sample size decreases. With $N < 50$, distances between subsample mean shapes and the observed mean are on average 1.5 to almost four times larger than the average distance between the latter and the average of bootstrap means (i.e., the uncertainty in the estimate of the observed mean). The variation in mean shapes is much smaller if compared to the interspecific distance between *C. aethiops* and *C. mitis*. However, with $N < 30$ the distance between subsample mean shapes and the observed mean can be on average up to 37% of the interspecific distance. Precision also deteriorates as sample size becomes smaller.

Angles between allometric trajectories (Table 3, Fig. 3a–b) quickly become much larger than the observed

Table 1 Size (mm) variation: observed, bootstrap and random subsamples

	<i>N</i>	Mean of CS	Percentile		SD of CS	Percentile	
			2.5th	97.5th		2.5th	97.5th
Females							
Observed	169	253.2	–	–	12.2	–	–
Bootstrap	169	253.1	251.5	255.3	12.1	11.0	13.0
Subsamples	100	253.2	251.8	254.3	12.2	11.3	13.0
	90	253.2	251.8	254.7	12.2	11.2	13.2
	80	253.1	251.5	254.7	12.2	11.2	13.2
	70	253.1	251.0	255.1	12.3	11.0	13.5
	60	253.0	250.4	255.3	12.0	10.8	13.5
	50	253.4	250.7	256.5	12.3	10.5	13.9
	40	253.2	250.3	256.5	12.1	10.1	14.4
	30	253.4	249.7	257.4	12.1	9.3	14.8
	20	253.6	249.5	258.6	12.1	8.6	15.1
	10	253.1	246.3	260.5	11.7	7.0	15.9
Males							
Observed	227	291.6	–	–	17.5	–	–
Bootstrap	227	291.4	289.4	293.5	17.4	16.1	18.8
Subsamples	100	291.8	289.1	294.4	17.5	15.9	19.0
	90	291.8	288.9	294.7	17.5	15.9	19.1
	80	291.8	287.6	295.2	17.5	15.6	19.6
	70	291.8	288.0	294.8	17.5	15.0	19.7
	60	291.7	288.3	294.7	17.7	15.4	19.9
	50	291.7	287.7	295.2	17.4	14.8	20.1
	40	291.5	285.7	296.1	17.5	14.8	21.4
	30	291.5	285.4	296.8	17.4	14.2	21.6
	20	291.4	283.7	298.0	17.2	12.4	22.4
	10	291.6	279.6	300.2	16.6	9.8	22.9

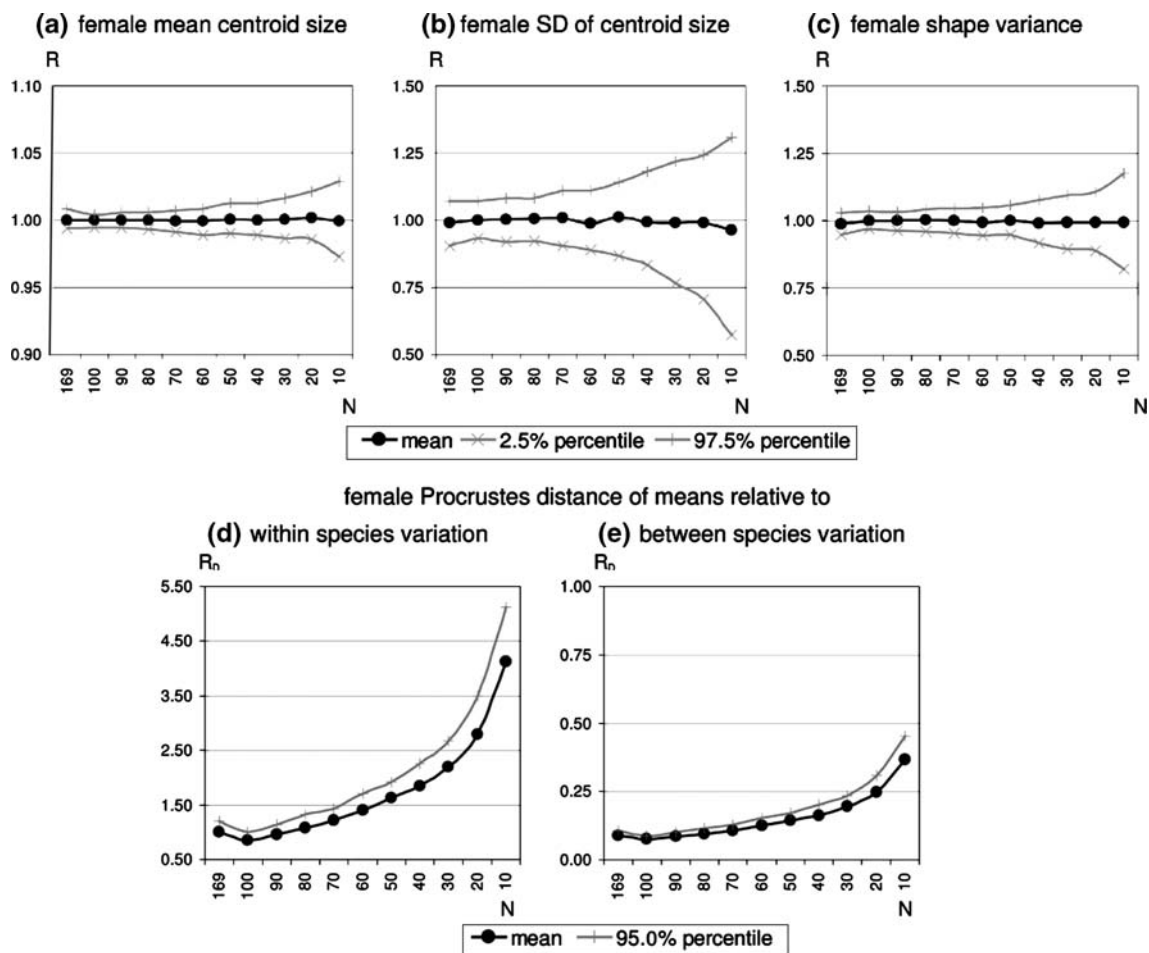


Fig. 2 a–e Profile plots of R and R_D for female means and variances: **a** mean and **b** standard deviation of size; **c** shape variance; sampling error in mean shapes relative to **d** within-species variation or **e** between species differences (*C. aethiops* to *C. mitis*). In all plots, the

average and percentiles of a parameter are first (left side) shown for the bootstrap samples and then for increasingly smaller subsamples ($N = 100, 90, \dots, 20, 10$). Please note that the vertical axis of profile plots may be drawn to different scales

when the sample size becomes smaller. This happens both in the comparison of static allometric trajectories within species (females and males of *C. aethiops*) as well as in the between species (*C. aethiops* versus *C. mitis*) comparison with pooled sexes. With sample sizes of each group ≤ 70 (within species) or ≤ 60 (interspecific), angles are on average 1.5 times larger than the observed and can be almost three times larger than that when only ten individuals are present in each group. Again, in addition to the increased inaccuracy of angle estimates, precision is apparently reduced in small samples.

The accuracy of the goodness of fit of different regression models seems less strongly affected by sampling error in small samples. This is found in both within (Table 3, Fig. 3c–d) and between (Table 3, Fig. 3e–g) species comparisons of static allometries. Only in the smallest samples does the fit of the regressions become close to or as large as 1.5 times the observed percentage. Also, although differ-

ences in percentages of explained variance of different regression models become larger in smaller samples, most of them tend to suggest similar patterns of relative magnitudes. Thus, within species (females versus males), the best-fit model with totally separate regressions explains less than 2% variance more than a single regression line fitting all specimens regardless of sex. This difference is still less than 4% in samples of just 30 specimens and only becomes much larger ($\approx 10\%$) in those of just ten individuals. Similarly, between species (*C. aethiops* versus *C. mitis*), lines with separate or parallel slopes explain about the same percentage of variance in the original sample but their fit is clearly larger than that of a single regression line. These relative differences remain largely unchanged even if the species' sample size is 20. Then, separate slopes explain about 2% more shape variance than parallel lines and each of them has a much larger fit ($>10\%$) than that of a single regression line for both species. Thus, comparisons

Table 2 Variation in Procrustes distances (PRD) of sample means to the observed mean, and in shape variance: observed, bootstrap and random subsamples

	<i>N</i>	Mean PRD	95th percentile	Mean shape variance	Percentiles	
					2.5th	97.5th
Females						
Observed	169	–	–	0.00348	–	–
Bootstrap	169	0.00438	0.00530	0.00345	0.00330	0.00359
Subsamples	100	0.00373	0.00441	0.00349	0.00337	0.00361
	90	0.00420	0.00501	0.00348	0.00336	0.00360
	80	0.00475	0.00580	0.00349	0.00334	0.00363
	70	0.00536	0.00630	0.00349	0.00333	0.00364
	60	0.00618	0.00751	0.00346	0.00329	0.00366
	50	0.00716	0.00841	0.00348	0.00330	0.00368
	40	0.00807	0.00991	0.00345	0.00319	0.00375
	30	0.00965	0.01171	0.00346	0.00311	0.00382
	20	0.01223	0.01531	0.00346	0.00309	0.00386
	10	0.01810	0.02242	0.00346	0.00286	0.00409
Males						
Observed	227	–	–	0.00345	–	–
Bootstrap	227	0.00383	0.00451	0.00341	0.00326	0.00355
Subsamples	100	0.00433	0.00531	0.00344	0.00330	0.00362
	90	0.00481	0.00591	0.00344	0.00330	0.00363
	80	0.00532	0.00651	0.00344	0.00327	0.00364
	70	0.00587	0.00701	0.00344	0.00325	0.00369
	60	0.00637	0.00781	0.00346	0.00323	0.00375
	50	0.00716	0.00841	0.00341	0.00318	0.00375
	40	0.00839	0.01021	0.00345	0.00309	0.00373
	30	0.00996	0.01221	0.00346	0.00309	0.00380
	20	0.01241	0.01502	0.00342	0.00300	0.00390
	10	0.01839	0.02263	0.00345	0.00289	0.00406

of static allometries between sexes of *C. aethiops* tend to suggest that females and males follow parallel and nearly overlapping trajectories, while comparisons of *C. aethiops* and *C. mitis* seem to indicate distinct but approximately parallel trajectories. The precision of estimates of the goodness of fit of the three models follows a trend similar to that described for the average estimate with a strong deterioration in the smallest samples (deterioration which is particularly visible in the within species comparison of female and male *C. aethiops*).

Discussion

Size parameters

The accuracy and precision of estimates of parameters that describe the variation of skull form in *C. aethiops* adults vary remarkably depending on the parameter. The mean size is very accurate and precise. A sample of just ten specimens allows estimates of size which are at worst

about 10 mm smaller or larger than the observed. This can be explained by looking at the standard deviation of size in the original sample relative to the mean size. The coefficient of variation ($100 \times \text{SD}/\text{mean}$) is small in both females (4.8%) and males (6.0%), despite having sampled specimens over the entire distribution range of *C. aethiops*, which covers most of Sub-Saharan Africa. The coefficient of variation of size of *C. aethiops* is not dissimilar to that of the skull of other cercopithecoid species (0.7–7.5%, A. Cardini and S. Elton submitted data) or even that of very distantly related mammals like marmots (0.5–4.0%, Cardini et al. 2007a). Thus, the result that a small sample size does not strongly affect estimates of mean size is likely to hold for many other mammal species.

The standard deviation of size in a sample is also accurate but not as precise as the mean. The variation in the standard deviation of size becomes large when less than 40 specimens are available. Thus, comparisons of size disparity to test whether size varies more in one population than in the other are better performed using relatively large samples ($N \geq 40$).

Table 3 Variation in angles (degrees) between static allometric vectors of group 1 (a, females; b, *C. aethiops*) and group 2 (a, males; b, *C. mitis*) and percentages of shape variance explained by size using different regression models (separate slopes, parallel lines, same line)

N		Vector angle			% Explained variance								
					Same line			Parallel lines			Separate slopes		
Group		Mean	Percentile		Mean	Percentile		Mean	Percentile		Mean	Percentile	
1	2		2.5th	97.5th		2.5th	97.5th		2.5th	97.5th		2.5th	97.5th
(a) Within-species (females versus males)													
Observed													
169	227	23.4	–	–	17.2	–	–	18.4	–	–	18.6	–	–
Bootstrap samples													
169	227	30.5	24.9	37.0	17.3	14.0	20.4	18.6	15.3	21.9	19.1	17.3	20.6
Random subsamples													
100	100	30.4	25.7	35.3	17.3	15.6	18.8	18.7	17.0	20.2	19.1	17.5	20.6
90	90	31.7	26.0	38.0	17.4	15.2	19.2	18.8	16.8	20.5	19.3	17.3	21.0
80	80	33.2	27.5	38.8	17.5	15.3	19.5	19.0	16.6	20.8	19.5	17.2	21.4
70	70	35.0	28.4	41.5	17.7	15.2	19.8	19.2	16.8	21.3	19.9	17.5	22.0
60	60	38.5	30.1	45.2	17.8	14.9	20.4	19.4	16.5	21.8	20.2	17.3	22.5
50	50	40.8	33.4	48.7	17.8	14.8	20.6	19.6	16.8	22.1	20.4	17.7	23.0
40	40	43.6	34.0	54.4	18.0	14.4	22.1	20.1	16.6	24.0	21.1	17.8	24.9
30	30	48.7	39.4	61.9	18.2	14.4	23.1	20.7	17.1	25.2	22.1	18.4	26.5
20	20	56.6	42.8	73.1	18.8	13.6	25.1	22.0	16.7	28.1	24.2	19.4	30.3
10	10	70.1	53.8	88.1	22.1	12.8	30.7	27.3	20.0	35.3	31.8	24.6	40.1
(b) Between species (<i>C. aethiops</i> versus <i>C. mitis</i>)													
Observed													
396	146	17.4	–	–	18.2	–	–	26.2	–	–	26.5	–	–
Bootstrap samples													
396	146	20.3	17.7	23.3	18.3	16.6	19.7	26.5	24.8	27.8	26.9	25.1	28.7
Random subsamples													
100	100	21.1	18.2	24.2	19.3	17.0	21.6	29.3	27.7	31.0	29.9	28.2	31.5
90	90	22.0	18.8	25.4	19.4	16.9	21.7	29.4	27.4	31.3	30.0	27.9	31.9
80	80	22.7	19.4	26.4	19.4	16.5	21.6	29.5	27.5	31.6	30.2	28.2	32.2
70	70	23.7	19.9	27.7	19.5	16.4	22.0	29.7	27.5	31.9	30.4	28.1	32.7
60	60	25.3	21.1	30.8	19.4	16.2	23.2	29.7	26.9	32.4	30.5	27.8	33.2
50	50	26.6	21.8	31.2	19.7	16.0	23.1	30.1	27.2	33.5	31.0	28.0	34.4
40	40	29.0	23.6	35.2	19.9	15.4	24.0	30.4	27.0	34.1	31.5	28.0	35.3
30	30	32.6	26.1	38.8	19.9	15.4	24.7	30.9	27.8	35.0	32.3	29.5	36.6
20	20	37.8	29.9	46.5	21.3	14.9	28.0	32.5	27.6	37.2	34.5	29.3	39.4
10	10	49.1	35.4	68.7	23.3	14.0	32.8	36.4	29.6	42.6	40.2	34.1	46.3

Shape parameters

Estimates of shape variance do remarkably well both in terms of accuracy and precision, both being more accurate and precise than are standard deviations of size. This suggests that, provided the specimens have been sampled over all or most of the distribution range of a species, relatively small samples ($N > 10$) could provide fairly accurate estimates of the magnitude of shape variation in a population of *C. aethiops*. The same cannot be said for

estimates of mean shapes. Their accuracy, compared to the uncertainty of the original sample mean shape, rapidly deteriorates when sample size decreases. Probably even more concerning is the fact that in samples of less than 30 specimens, the error in the mean shape estimate can be on average as large as 20–37% of the interspecific distance between mean shapes of *C. aethiops* and *C. mitis*, two species that diverged about 8 million years ago (Tosi et al. 2005) and which have profound differences in their ecology and behavior (the former being terrestrial and the latter

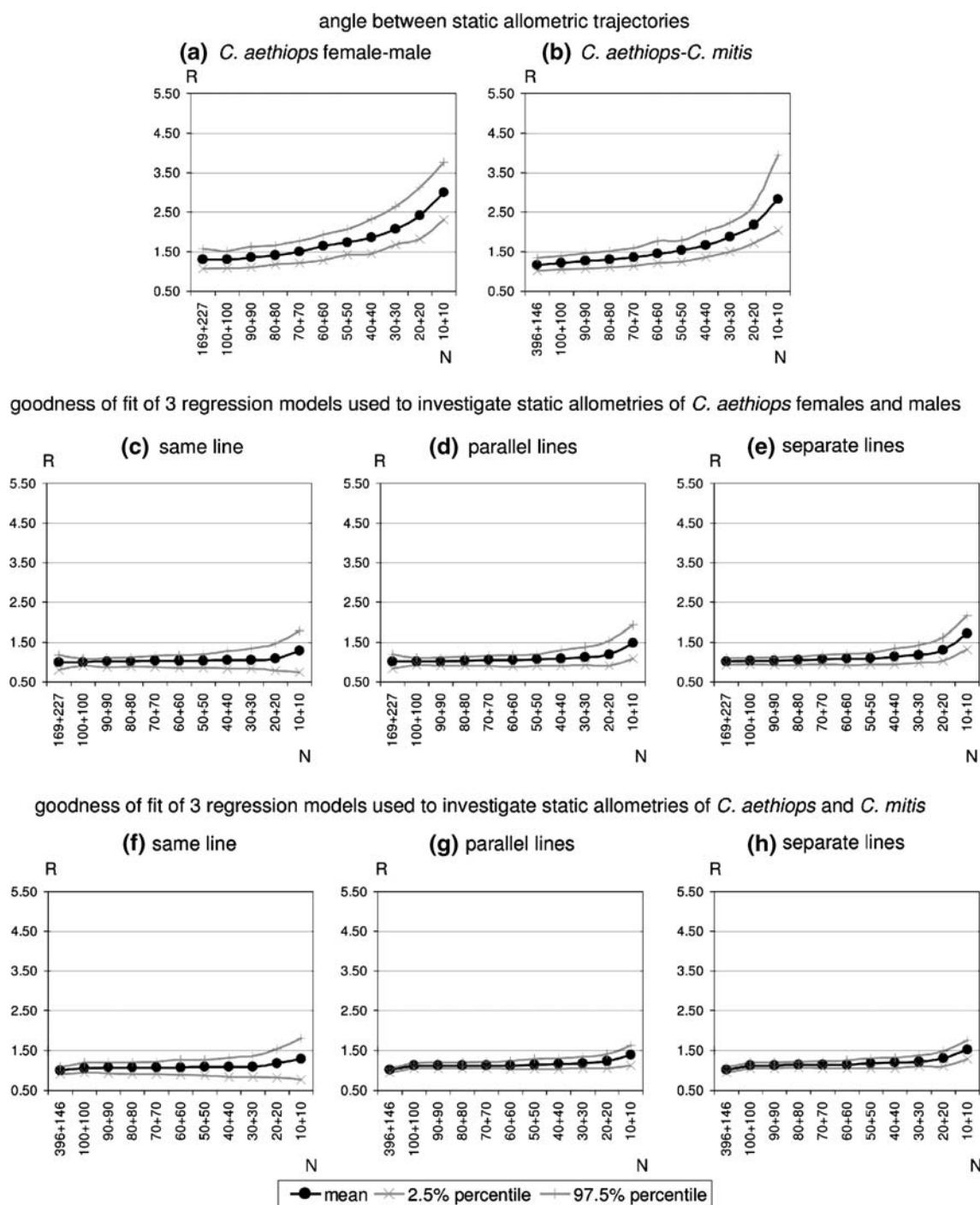


Fig. 3 a–f Profile plots of R for trajectory parameters: **a** angles between female and male static allometric trajectories; **b** angles between *C. aethiops* and *C. mitis* static allometric trajectories; percentages of variance explained by different regression models in the within species, **c–e** comparison and in the between species, **f–h**

comparison. In all these plots, N refers to number of specimens in the two groups whose trajectories are compared (for instance, $N = 169 + 227$ means that bootstrap samples included 169 females and 227 males, $N = 100 + 100$ means that subsamples included 100 females and 100 males, etc.)

arboreal). A few examples of recent geometric morphometric analyses that have investigated various aspects of morphological variation in clades of mammals are listed in Table 4 together with the minimum and average species

sample size (with pooled sexes) from those studies. Several of them include species with divergence times comparable to that of *C. aethiops* and *C. mitis*. About 50% of these studies have an average sample size smaller than 30 and

Table 4 Mean and minimum sample size of species in some recent geometric morphometric studies on mammals (ordered according to increasing mean *N*)

Reference	Taxon	Structure	Type of study	Number of species	<i>N</i>	
					Min	Mean
Bruner et al. (2003)	Humans	Cranium	Palaeontology	3 ^a	7	8.3
Milne and O'Higgins (2002)	Kangaroos	Cranium	Form variation	7	1	9.3
Nicola et al. (2003)	Spiny rats	Cranium	Form variation	5	8	13.2
Rohlf et al. (1996)	Moles	Cranium	Form variation	7	3	16.1
Pan et al. (2003)	Macaques	Cranium	Form variation	5	6	18.4
Caumul and Polly (2005)	Marmots	Skull	Ecomorphology	5	3	20.2
Ponce de León and Zollikofer (2001)	Humans	Skull	Palaeontology	2	16	20.5
Dobigny et al. (2002)	Gerbils	Cranium	Form variation	4	1	20.8
Cardini (2003)	Marmots	Mandible	Form variation	14	3	27.7
Singleton (2002)	Papionins	Cranium	Ontogeny	8	15	29.6
Singleton (2005)	Cercopithecids	Cranium	Form variation	15	6	30.0
Collard and O'Higgins (2001)	Papionin	Cranium	Ontogeny	7	8	31.7
Delson et al. (2001)	Humans	Cranium	Palaeontology	3	2	41.3
Leigh (2006)	Papionin	Cranium	Ontogeny	5 ^a	11	42.8
Penin et al. (2002)	Apes	Cranium	Ontogeny	2	41	45.5
Frost et al. (2003)	Baboons	Cranium	Ecomorphology	9 ^a	15	50.2
Mitteroecker et al. (2004)	Apes	Cranium	Ontogeny	5	?	53.6
A. Cardini and S. Elton (submitted data)	Guenons	Skull	Form variation	23	5	57.2
Harvati (2003)	Humans	Cranium	Palaeontology	5 ^a	2	58.0
Cardini and Thorington Jr (2006)	Marmots	Cranium	Ontogeny	6	21	67.5

^a Including subspecies

most of them include at least one species with a sample of less than ten individuals.

If confirmed in other groups, findings about the differing effects of sampling error on estimates of species means and variances of size and shape have implications for taxonomic comparisons of closely related species (for instance, at the level of tribe). Reconstructions of phenetic or phyletic relationships based on species mean shapes can be strongly affected by errors in small samples. In the worst case, very small samples could pick up highly divergent phenotypes that represent local transient variation unlikely to be fixed in the population gene pool (Polly, in press), but capable of introducing a strong bias in the inferred pattern of interspecific relationships. Indeed, it is not unusual for studies of museum specimens to have relatively few representatives of rare species that are often collected at the same time and in the same locality. Similar, and often more serious, issues can arise in palaeontological studies, where closely related taxa with few and often fragmentary specimens are compared.

Errors in sample mean shapes can also have profound consequences on studies of disparity (Foote 1997) in recent clades. Partial disparities, for instance, are computed using distances between species means and the grand mean of all species (Zelditch et al. 2004), and errors are thus likely to

be large in small samples. Standard errors in measures of partial disparity estimated with bootstraps (Zelditch et al. 2004) can help to understand uncertainties of disparity estimates and to decide if and what species might be excluded from the analysis to improve accuracy. A. Cardini and S. Elton (submitted data) found significant negative correlations ($r = -0.441$, $P < 0.05$) between standard errors of shape partial disparities and sample size in a geometric morphometric analysis of the skull of 21 species of guenons. That this observation may hold in other taxa is suggested by a similar negative correlation ($r = -0.566$, $P < 0.05$) between standard errors of partial disparity and sample size found by Cardini et al. (2007a) in a study of cranial variation in 13 species of marmots.

'Size versus shape'

Interestingly, the effect of sampling error on sample means and variances (or standard deviations) is not the same on size as it is on shape. Estimates of mean shapes are not very reliable in small samples, while those of mean size always tend to be accurate and precise. In contrast, standard deviations of size vary clearly more than do shape variances when samples are small, although both parameters tend to be accurate. This suggests that

even if estimates of sample mean shapes can be far from the population mean in small samples, the variation around these means will not be very dissimilar to that observed in larger samples and, presumably, will provide a relatively valid estimate of the population variance. If this observation can be generalized, then comparisons of within species shape disparity (i.e., population variances) can be fairly accurate even when relatively few specimens are available. In contrast, an accurate comparison of shape partial disparity within a clade of closely related species using mean shapes will require relatively large samples. Also, the fact that a smaller sample size has less effect on shape variance than on mean shape estimates might have implications on type I error rates, possibly leading to a number of rejections of null hypotheses greater than the nominal alpha.

Regression parameters

The accuracy and precision of estimates of angles between static allometric trajectories deteriorate rapidly as sample size becomes smaller both in the within and between species comparisons. Angles tend to be strongly overestimated in small samples. This also explains why angles in bootstrap samples, where the sample size is constant but some specimens are present more than once in the same sample (thus, reducing the effective sample size), tend to be larger than the observed. Indeed, static allometries in the absence of a large size variation tend to be short and with a large scatter around the regression line, which is the same as saying that the variation is almost isotropic. This is particularly evident for the within species comparison (results not shown) where a small (<8.6%) but significant percentage of variance is explained by the regression of shape onto size within each sex. However, even when a larger shape variation is found in the direction of size, like in the interspecific comparison of *C. aethiops* (percentage of variance explained by the regression of shape onto size, 17.3%) and *C. mitis* (19.8%), angles are more than 50% larger than the observed when specimens are 50 or less. Thus, estimates of regression coefficients become easily affected by the sample composition and these errors will, in turn, lead to inaccuracies in estimates of angles between group trajectories. This observation is supported by A. Cardini and S. Elton's (submitted data) comparison of static allometric trajectories of guenons in which average pairwise angles of species trajectories are negatively correlated with sample size according to a power relationship. Thus, comparison of angles of static allometries should be performed with great caution and using large samples even when size-related shape variation is relatively large, as is the case in the between species comparison of

C. aethiops and *C. mitis*. Predictions of shapes along trajectories should also be done with great care, especially if differences in shapes at the extremes of a trajectory are magnified, as it is often done to help visualization.

Ontogenetic trajectories are longer than static allometries and thus less likely to be very strongly affected by sampling error. Cobb and O'Higgins (2004) did a randomized selection experiment similar to the one reported here to test the effect of sample size on estimates of angles between ontogenetic trajectories of *P. troglodytes* and *P. paniscus* and found that at least 15–20 individuals are needed to get relatively accurate estimates. This 'minimum desirable' sample size is indeed much smaller than the about 60 specimens of our comparison of interspecific static allometries. However, it is likely that since Cobb and O'Higgins (2004) did the randomization and selection in only one of the two samples, with the other sample always including all the available specimens, their analysis underestimates the 'minimum desirable' sample size for accurate estimates of angles between ontogenetic trajectories.

Percentages of variance explained by regressions are generally accurate and fairly precise, even in small samples. It is unsurprising that the direction of trajectories, and thus the angles between them, may be highly variable in small samples, while the fit of regressions does not change very much when the shape variation which is explained by the independent variable is not large. Differences in a few specimens might easily change the slope of the regression line, but the scatter around it, and thus the fit of the model, will be more or less the same. To detect a small signal, large samples are needed both for testing significance and for inferring population parameters. It is, however, interesting to note that differences in the fit of the three regression models in subsamples tend to be consistent with the observed and are not strongly affected by sample size. The ratio between percentages of allometric shape of different models can be used to measure how much better a model fits compared to another one. This ratio is equal to 0.99 in the comparison of separate slopes to parallel lines and 0.93 in the comparison of parallel lines to a single line in the original samples of females and males of *C. aethiops*. These ratios are still remarkably close, on average, to the observed when samples of just 30 specimens are compared, being 0.88 and 0.94, respectively. Similarly, in the between species comparison, the observed ratios are (separate/parallel) 0.99 and (parallel/single) 0.69 and the average ratios in subsamples of 20 specimens are, respectively, 0.94 and 0.66. Thus, it might be possible to get some indication of models of trajectory divergence even in relatively small samples by comparing their goodness of fit.

Conclusions

This study has focused on the effect of sampling error on estimates of population parameters in geometric morphometrics of representatives of cercopithecine monkeys. These parameters can have a purely descriptive function, like the average size of the skull in a species. However, they are also often employed in statistical tests, which will thus themselves be more or less strongly affected by sampling error. Means and variances of size and shape are used to test group differences in analyses of variances and discriminant analyses or to summarize group relationships with ordinations and cluster analyses (examples can be found in the references reported in Table 4, and in Zelditch et al. 2004). These parameters are also used in analyses of morphological disparity (Foote 1997; Zelditch et al. 2004), which investigate the amount of form diversity within clades. Angles between trajectories are measured to estimate the amount of divergence in shape variation associated with other factors, including size (allometry), time (development), ecology and geography (clinal variation). Permutation tests can be used to test whether angles are significantly larger than zero. Different models of regression can also be used to test for the significance of differences in slope (separate versus parallel) and intercept (parallel versus single line) in trajectories. Analyses of the divergence in trajectories are especially common in ontogenetic studies (examples in Table 4), where changes that take place during ontogeny are compared across groups to understand whether patterns are the same and differences arise early in development (parallel lines) or whether differences are produced by simple extension/truncation of a common trajectory (overlapping lines).

Addressing all the issues mentioned above is beyond the scope of this paper. Instead, the goal here has been to understand how quickly and how strongly estimates of different parameters can become unreliable when sample size decreases. The empirical approach adopted in our study illustrates how easily clues about the likely minimum desirable sample size in geometric morphometric studies (that investigate aspects of morphological variation using means of size and shape variables, their variances and so on) can be obtained by simple repeated randomized selection experiments. This kind of rarefaction analysis can be directly performed on the study group or using a sample taken from a closely related taxon as a proxy, as might be desirable in palaeontology, for instance, where samples of fossil mammals are generally very small. The empirical approach has another advantage, besides its simplicity. It avoids the need to test how closely the real data are to the assumptions of simulations, a task which is often difficult to achieve, thus making it hard to understand whether

generalizations from mathematical modeling hold in the specific dataset under study.

In summary, the increasingly widespread application of geometric morphometrics requires a better understanding of how statistical and visual results of often sophisticated analyses may be strongly affected by a simple factor: inadequate sampling. The centrality of this apparently neglected issue does not only concern the statistical description and comparison of size and shape, but also the last stage of a geometric morphometric study, the visualization of shape variation. Shapes along a trajectory are often used to describe the variation associated with a specific factor and clues on biological processes are often inferred by the inspection of shapes predicted by regressions. Variation of predicted shapes is frequently magnified to make differences more evident. However, these predicted shapes (as well as the mean shapes of groups) may be not be representative of real biological variation, and interpretations based on them may be misleading.

Acknowledgments We are greatly in debt to all museum curators and collection managers who gave us permission to study the specimens in their care and helped during data collection. A large number of people have contributed in different ways to this study. We are deeply grateful to all of them, and apologize to those whom we might have forgotten to acknowledge. Special thanks are due to P. O'Higgins, University of York; F. J. Rohlf, State University of New York; C. P. Klingenberg, University of Manchester, Manchester; P. D. Polly, Indiana University, Bloomington. We also very much appreciate the comments of an anonymous referee and the Editor in improving this work, and thank them for their time and trouble. This study was funded by a grant from the Leverhulme Trust.

References

- Adams DC, Slice DE, Rohlf FJ (2004) Geometric morphometrics: ten years of progress following the 'revolution'. *Ital J Zool* 71:5–16
- Bookstein FL (1991) *Morphometric tools for landmark data*. Cambridge University Press, Cambridge, MA
- Bruner E, Manzi G, Arsuaga JL (2003) Encephalization and allometric trajectories in the genus *Homo*: evidence from the Neandertal and modern lineages. *Proc Natl Acad Sci USA* 100:15335–15340
- Cardini A (2003) The geometry of marmot (Rodentia: Sciuridae) mandible: phylogeny and patterns of morphological evolution. *Syst Biol* 52:186–205
- Cardini A, Thorington RW Jr (2006) Post-natal ontogeny of the marmot (Rodentia, Sciuridae) cranium: allometric trajectories and species divergence. *J Mammal* 87:201–216
- Cardini A, Thorington RW Jr, Polly PD (2007a) Evolutionary acceleration in the most endangered mammal of Canada: phylogenetic signal and cranial divergence in the Vancouver Island marmot (Rodentia, Sciuridae). *J Evol Biol* (in press). doi:10.1111/j.1420-9101.2007.01398.x
- Cardini A, Jansson A-U, Elton S (2007b) Ecomorphology of vervet monkeys: a geometric morphometric approach to the study of clinal variation. *J Biogeogr* (in press). doi:10.1111/j.1365-2699.2007.01731.x

- Caumul R, Polly PD (2005) Phylogenetic and environmental components of morphological variation: skull, mandible and molar shape in marmots (*Marmota*, Rodentia). *Evolution* 59:2460–2472
- Collard M, O'Higgins P (2001) Ontogeny and homoplasy in the papionin monkey face. *Evol Dev* 3:322–331
- Cobb SN, O'Higgins P (2004) Hominins do not share a common postnatal facial ontogenetic shape trajectory. *J Exp Zool B* 302:302–321
- Delson E, Harvati K, Reddy D, Marcus LF, Mowbray K, Sawyer GJ, Jacob T, Márquez S (2001) The Sambungmacan 3 *Homo erectus* calvaria: a comparative morphometric and morphological analysis. *Anat Rec* 262:360–377
- Dobigny G, Baylac M, Denys C (2002) Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* (Rodentia, Gerbillinae) species. *Biol J Linn Soc* 77:319–327
- Dryden IL, Mardia KV (1993) Multivariate shape analysis. *Sankhya* 55(A):460–480
- Dryden IL, Mardia KV (1998) Statistical shape analysis. Wiley, New York
- Foote M (1997) The evolution of morphological diversity. *Ann Rev Ecol Syst* 28:129–152
- Frost SR, Marcus LF, Bookstein FL, Reddy DP, Delson E (2003) Cranial allometry, phylogeography, and systematics of large-bodied papionins (primates: cercopithecinae) inferred from geometric morphometric analysis of landmark data. *Anat Rec* 275A:1048–1072
- Harvati K (2003) Quantitative analysis of Neanderthal temporal bone morphology using 3-D geometric morphometrics. *Am J Phys Anthropol* 120:323–338
- Klingenberg CP, Monteiro LR (2005) Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Syst Biol* 54:678–688
- Leigh SR (2006) Cranial ontogeny of *Papio* baboons (*Papio hamadryas*). *Am J Phys Anthropol* 130:71–84
- Milne N, O'Higgins P (2002) Inter-specific variation in *Macropus* crania: form, function and phylogeny. *J Zool* 256:523–535
- Mitteroecker P, Gunz P, Bernhard M, Schaefer K, Bookstein FL (2004) Comparison of cranial ontogenetic trajectories among great apes and humans. *J Hum Evol* 46:679–697
- Nicola PA, Monteiro LR, Pessoa LM, Von Zuben FJ, Rohlf FJ, Dos Reis SF (2003) Congruence of hierarchical, localized variation in cranial shape and molecular phylogenetic structure in spiny rats, genus *Trinomys* (Rodentia: Echimyidae). *Biol J Linn Soc* 80:385–396
- O'Higgins P (1999) Ontogeny and phylogeny: morphometric approaches to the study of skeletal growth and evolution. In: Chaplain MAJ, Singh GD, McLachlan J (eds) *On growth and form: Spatio-temporal patterning in biology*. Wiley, New York, pp 373–393
- O'Higgins P, Jones N (2006) Morphologika 2.2. Tools for shape analysis. Hull York Medical School, University of York, York. <http://www.york.ac.uk/res/fme/>
- Pan R, Wei F, Li M (2003) Craniofacial variation of the Chinese macaques explored with Morphologika. *J Morphol* 256:342–348
- Penin X, Berge C, Baylac M (2002) Ontogenetic study of the skull in modern humans and common chimpanzees: neotenic hypothesis reconsidered with a tridimensional Procrustes analysis. *Am J Phys Anthropol* 118:50–62
- Polly PD (2005) Development and phenotypic correlations: the evolution of tooth shape in *Sorex araneus*. *Evol Dev* 7:29–41
- Polly PD Phylogeographic differentiation in *Sorex araneus*: is the species morphologically homogenous? *Russian Ther J* (in press)
- Ponce de León MS, Zollikofer CPE (2001) Neanderthal cranial ontogeny and its implications for late hominid diversity. *Nature* 412:534–538
- Rohlf FJ (1998) On applications of geometric morphometrics to study of ontogeny and phylogeny. *Syst Biol* 47:147–158
- Rohlf FJ (2000a) On the use of shape spaces to compare morphometric methods. *Hystrix It J Mamm* 11:8–24
- Rohlf FJ (2000b) Statistical power comparisons among alternative morphometric methods. *Am J Phys Anthropol* 111:463–478
- Rohlf FJ (2003a) Bias and error in estimates of mean shape in morphometrics. *J f Evol* 44:665–683
- Rohlf FJ (2003b) TpsSmall. Department of Ecology and Evolution, State University of New York, Stony Brook, New York. <http://life.bio.sunysb.edu/morph/>
- Rohlf FJ (2005) NTSYS-pc, version 2.20d. Exeter Software, Setauket, New York
- Rohlf FJ, Marcus LF (1993) A revolution in morphometrics. *Trends Ecol Evol* 8:129–132
- Rohlf FJ, Slice DE (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst Zool* 39:40–59
- Rohlf FJ, Loy A, Corti M (1996) Morphometric analysis of old world Talpidae (Mammalia, Insectivora) using partial-warp scores. *Syst Biol* 45:344–362
- Singleton M (2002) Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). *J Hum Evol* 42:547–578
- Singleton M (2005) Craniofacial shape variation in cercopithecine primates. In: Slice D (ed) *Modern morphometrics in physical anthropology*. Kluwer, New York, pp 313–347
- Slice DE (1999) Morpheus, beta version. Department of Ecology and Evolution, State University of New York, Stony Brook, New York
- Slice DE (2005) Modern morphometrics. In: Slice D (ed) *Modern morphometrics in physical anthropology*. Kluwer, New York, pp 1–45
- SPSS for Windows (2004) SPSS Inc., Version 11.5.0 SPSS Inc., Chicago, IL, USA
- Tosi AJ, Detwiler KM, Disotell TR (2005) X-chromosomal window into the evolutionary history of the guenons (Primates: Cercopithecini). *Mol Phylogenet Evol* 36:58–66
- Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) *Geometric morphometrics for biologists: a primer*. Elsevier Academic, Amsterdam