View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Archivio istituzionale della ricerca - Università di Modena e Reggio Emilia

Biological Journal of the Linnean Society, 2008, **93**, 813–834. With 6 figures

Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons

ANDREA CARDINI[1,2]* and SARAH ELTON[2]

[1]*Museo di Palaeobiologia e dell'Orto Botanico, Universitá di Modena e Reggio Emilia, via Università 4, 41100, Modena, Italy*
[2]*Hull York Medical School, The University of Hull, Cottingham Road, Hull HU6 7RX, UK*

Received 15 August 2007; accepted for publication 11 October 2007

Form and genes often tell different stories about the evolution of animals, with molecular data generally considered to be more objective than morphological data. However, form provides the basis for the description of organisms, and the study of fossils crucially depends on morphology. Complex organisms tend to evolve as 'mosaics', in which parts may be modified at varying rates and in response to different selective pressures. Thus, individual anatomical regions may contain different phylogenetic signals. In the present study, we used computerized methods to 'dissect' the skulls of a primate clade, the guenons, into functional and developmental modules (FDM). The potential of different modules as proxies for phylogenetic divergence in modern lineages was investigated. We found that the chondrocranium was the only FDM in which shape consistently had a strong and significant phylogenetic signal. This region might be less susceptible to epigenetic factors and thus more informative about phylogeny. The examination of the topology of trees from the chondrocranium suggested that the main differences evolved at the time of the radiation of terrestrial and arboreal guenons. However, phylogenetic reconstructions were found to be strongly affected by sampling error, with more localized anatomical regions (i.e. smaller/less complex FDMs) generally producing less reproducible tree topologies. This finding, if confirmed in other groups, implies that the utility of specific FDMs for phylogenetic inference could, in many cases, be hampered by the low reproducibility of results. The study also suggested that uncertainties due to sampling error may be larger than those from character sampling. This might have implications for phylogenetic analyses, which typically provide estimates of support of tree nodes based on characters but do not generally take into account the effect of sampling error on the tree topology. Nonetheless, studies of the potential of different FDMs as proxies for phylogenetic divergence in modern lineages, such as the present study, provide a framework that may help in modelling the morphological evolution of present and fossil species. © 2008 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2008, **93**, 813–834.

ADDITIONAL KEYWORDS: Cercopithecini – geometric morphometrics – morphological evolution – reliability and reproducibility – skull modules.

INTRODUCTION

Morphology provides the basis for systematic descriptions (MacLeod, 2002). In the vast majority of systematic studies, organisms are first grouped based on their morphology and then compared to infer their relationships. This is also the starting point in molecular phylogenetics (Jensen, 2003). The analysis of morphology is thus crucial to the study of phylogeny in both modern and ancient organisms (MacLeod, 2002). However, since the advent of molecular biology and its exponentially increasing application to phylogenetic reconstruction, incongruencies between molecular and morphological phylogenies have frequently raised doubts about the possibility of recovering a phylogenetic signal in organismal form (Milinkovitch & Thewissen, 1997; Collard & Wood, 2000; Madsen *et al*., 2001; Gaubert & Veron, 2003; Koepfli *et al*., 2006).

Examples of the dichotomy between form and molecules in phylogenetic inference can be found at every

*Corresponding author. E-mail: alcardini@interfree.it; cardini@unimo.it

level of the taxonomic hierarchy. For example, strong molecular evidence from multiple genetic markers led to a profound reorganization of the metazoan classification, with the creation of two new major clades of Protostomia, the Ecdisozoa and Lophotrochozoa, and the disappearance of traditional groupings such as the Articulata, which consisted of the two major clades of segmented animals, the annelids and the arthropods (Halanych, 2004). Among mammals, multiple mitochondrial and nuclear DNA sequences indicate two well-separated clades, the Afrotheria and the Laurasiatheria, that had not been found in previous analyses based on morphology (Springer *et al*., 1997; Stanhope *et al*., 1998; Madsen *et al*., 2001; Murphy *et al*., 2001; Kriegs *et al*., 2006). Even among our closest relatives, the primates, morphology and molecules are not always in good agreement. In African apes, the study of hard tissue morphology has generally indicated a close relationship between chimps and gorillas, whereas molecular data overwhelmingly support a human–chimp clade (Pilbeam, 2000). Similarly, the 'long-faced' baboons and mandrills closely resemble each other, especially compared to mangabeys, but molecular studies consistently indicate a sister group relationship between baboons and black mangabeys on the one hand, and mandrills and capped mangabeys on the other (Disotell, 2000).

Molecular data are often considered to be more objective and rigorous than morphological data. DNA sequences contain four easily identified and mutually exclusive character states (Halanych, 2004). In addition, genes distant enough to segregate independently provide gene trees that are independent assays of evolutionary relationships (Pilbeam, 2000). Morphological characters are less easily defined and scored, and may belong to highly integrated structures, thus violating assumptions of independence in parsimony analyses. Regardless of whether or not this view is universally shared, the rapid increase in the amount of genetic information available for phylogenetic reconstruction and the improvement in the understanding of molecular processes suggest that DNA analysis will yield most of the answers about relationships among modern species.

This does not diminish the actual and potential utility of morphological characters for phylogenetic inference. By comparing different types of data, we can gain a better comprehension of how organisms evolve and understand what characters might be more conservative. In turn, this could help to elucidate which characters are reliable indicators of phylogenetic relationships and which might tend to reflect ecological adaptation. Also, with the exception of very recent fossils (or subfossils), molecules are unlikely to be of direct help to palaeontologists in the reconstruction of the phylogeny of extinct organisms. Thus, whether a phylogenetic signal can be recovered from morphology remains a central issue in palaeontology as well as more generally in evolutionary biology. Frequent investigations of this have been undertaken by cladists, through comparison of tree topologies from parsimony analyses of meristic morphological characters with molecular cladograms. However, morphological differences often occur across a continuous range of variation. Thus, measurements rather than character states need to be compared. This is the realm of morphometricians, many of whom are also interested in detecting the phylogenetic signal in their data. This notwithstanding, the majority of morphometric analyses are not aimed at directly inferring phylogenies. Instead, many morphometric descriptors are included and multivariate statistics used to test group differences and reconstruct similarity (phenetic) patterns, which are then interpreted on the basis of available phylogenies from independent datasets.

Landmark based morphometric methods for the geometric comparison of biological forms have been developed in the past two decades (Rohlf & Marcus, 1993; Adams, Slice & Rohlf, 2004; Slice, 2005) and have proved to be very powerful. One application has been the study of the phylogenetic relationships of modern and fossil species, with several authors suggesting that the greater accuracy and statistical power of geometric morphometric methods makes them particularly suitable for this task (Polly, 2001; Lockwood, Kimbel & Lynch, 2004). Often, the results of geometric morphometric analyses, usually 'phenetic' in scope, are discussed in the context of known (usually molecular) phylogenies (Fadda & Corti, 2001; Rüber & Adams, 2001; Milne & O'Higgins, 2002; Singleton, 2002; Cardini, 2003; Guy *et al*., 2003; Cardini & O'Higgins, 2004; Cardini & O'Higgins, 2004; Lockwood *et al*., 2004).

Less frequently, researchers have attempted to use geometric morphometrics to provide estimates of the magnitude of the phylogenetic signal in morphological data. Cole, Lele & Richtsmeier (2002) developed a bootstrap approach for measuring the uncertainty in morphometric trees compared with an available phylogeny. Using a relatively simple case study, a comparison of the facial skeleton of the four living genera of ateline monkeys, they found low congruence between the phenogram generated from facial measurements and a cladogram based on a variety of data (including genetics and behaviour), which they interpreted in terms of different rates of morphological evolution because of dietary adaptations. Caumul & Polly (2005) quantified the environmental and phylogenetic components of shape variation in crania, mandibles, and teeth of marmots using path analysis. A phylogenetic signal was recovered in all structures

(5–15% of total variance), with crania and molars found to be the best predictors of phylogenetic relationships. They argued that the good phylogenetic performance of the cranium may be explained by its structural complexity, whereas the poor results obtained with the mandible might be related to the confounding effect of its greater ecophenotypic plasticity. Other studies used matrix correlations to provide a more direct way of estimating the strength of the phylogenetic signal. Polly (2001, 2003), Nicola *et al*. (2003), Monteiro & Dos Reis (2005), and Macholán (2006) measured the correlation between matrices of shape distances and matrices of genetic distances; high correlations implied that the relative positions of species in the morphospace mirrored the pattern of (phylo-)genetic divergence, with significance calculated using permutation tests. Polly (2001, 2003) found a significant correlation between molar shape and divergence in cytochrome *b* gene sequences in both shrews and marmots, suggesting the potential of dental morphology as a proxy for phylogenetic divergence in phylogeographical reconstruction. Similarily, a large and significant correlation between genetic and morphological distances was found by Macholán (2006) in a study of molar shape variation in mice. In these studies, low environmental variance and high heritability might be an explanation for the effectiveness of teeth in recovering phylogenetic groups (Caumul & Polly, 2006). By contrast, Nicola *et al*. (2003) and Monteiro & Dos Reis (2005), in studies of cranial and mandibular variation in spiny rats, both failed to detect a strong phylogenetic signal. Monteiro & Dos Reis (2005) interpreted this outcome as a consequence of the complexity of interactions, morphogenetic rules, ecological phenomena, and stochastic or deterministic evolutionary forces that determine the overall shape of the mandible.

Such quantitative comparisons of shape differentiation and genetic divergence represent a fundamental step in understanding the dynamics of morphological evolution. By measuring the congruence between shape and gene differences, and by comparing the strength of the phylogenetic signal in different structures or among different groups, we can look for patterns that might be explained in terms of development or adaptation.

One stimulating area of potential research is the study of mosaic evolution in a phylogenetic context (Cole *et al*., 2002) because, within a group of related organisms, different structures or character complexes may have evolved under different scenarios (Cole *et al*., 2002; Polly, 2008). However, despite a great interest in the modularity and integration of complex structures such as the mammal cranium (Goswami, 2006) and the rapid adoption of geometric morphometrics in evolutionary developmental biology (evo-devo), little work appears to have been done in the context of phylogenetic inference.

In some cladistic analyses, anatomical structures have been split into regions to test their performance in phylogenetic reconstructions. For example, Collard & Wood (2001) examined the ability of four regional skull character groups to recover the phylogeny of either hominoids or papionins. Linear measurements were coded into meristic characters, which were subjected to parsimony analysis. When the resulting morphological trees were compared with molecular cladograms, it was found that all sets of morphological characters were more or less equally unreliable for phylogenetic inference. This type of approach to the 'modular study' of the phylogenetic signal in morphological data is valuable but has some limitations when applied to the shape variables generated using geometric morphometrics. In particular, information is inevitably lost when continuous variation is coded into discrete character states (Caumul & Polly, 2006). By contrast, methods for phylogenetic inferences that can be directly applied to shape variables can be used to demonstrate the potential of a geometric morphometric approach when quantifying the phylogenetic signal in different anatomical regions (Caumul & Polly, 2006).

The present study, with a focus on one clade of mammals, the guenons (Primates, Cercopithecini), is the first to combine a 'hierarchical modular approach' similar to that used in evo-devo research on integration and modularity (Cheverud, 1995, 1996; Ackermann & Cheverud, 2000; Hallgímsson *et al*., 2004) with geometric morphometric data to investigate how much information on phylogeny is stored generally in the skull and more specifically in its functional and developmental modules (FDMs). A fundamental assumption in virtually all previous analyses that use geometric morphometrics within a phylogenetic framework is that a pre-existing phylogeny is available for the study taxa and that this estimate has been made without error. Thus, the study taxon was chosen not only because it is highly diverse (for ecology, size variation, number of species and distribution), but also because its phylogenetic relationships were clarified by a recent phylogenetic analysis on a large amount of molecular data (Tosi, Detwiler & Disotell, 2005). The analysis by Tosi *et al*. (2005) is not only one of the most recent studies of guenon evolutionary relationships, but also is very comprehensive, sampling nearly all living guenon species. It is used here not as an 'absolute truth' but as the best proxy available to date, and one that for the most part mirrors previous phylogenetic analyses of the clade using chromosomal, morphological, behavioural, and acoustic data (Gautier, Vercauteren Drubbel & Deleporte, 2002).

Table 1. Species samples

Genus	Species	Taxonomic authority	Females	Males
Allenopithecus	*nigroviridis*	Lang 1923	7	8
Cercopithecus	*aethiops*	(Linnaeus, 1758)	169	227
	ascanius	(Audebert, 1799)	37	39
	cephus	(Linnaeus, 1758)	29	29
	diana	(Linnaeus, 1758)	32	32
	hamlyni	Pocock, 1907	15	15
	lhoesti	Sclater, 1899	16	18
	mitis	Wolf, 1822	67	79
	mona	(von Schreber, 1775)	16	19
	neglectus	Schlegel, 1876	24	27
	nictitans	(Linnaeus, 1758)	24	23
	petaurista	(von Schreber, 1774)	16	25
	pogonias	Bennett, 1833	38	38
Erythrocebus	*patas*	(von Schreber, 1774)	9	21
Miopithecus	*ogouensis*	Kingdon, 1997	16	11
Total			515	611

FDMs were defined on the basis of structural identity (cranium versus mandible), mode of ossification (chondrocranium versus dermatocranium), and main sources of epigenetic influences (brain growth, masticatory muscles, teeth); 'epigenetics' is used here following the definition of Herring (1993: 472): '*Sensu lato*, epigenetics refers to the dynamic interaction between the genome and its environment and its study is the study of the mechanisms which effect ontogeny'. The FDMs used in the present study have also been employed in previous studies (Cheverud, 1995) and are applied here as a priori modules to assess whether any have a correlation with the molecular phylogeny that is greater than would be expected by chance alone. Thus, the first main goal of the study was to test the significance of the phylogenetic signal of FDMs and compare its magnitude across sets of FDMs. The second was to provide estimates of the reproducibility of phenetic and phylogenetic reconstructions of interspecific relationships based on FDMs, and to discuss scenarios of morphological evolution suggested by tree topologies.

MATERIAL AND METHODS

Sample and data collection

The sample comprised 1126 adult specimens of 15 species of Cercopithecini (Table 1). 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 Harvard University (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), and 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). 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 Figure 1. Definitions of landmarks are provided in Cardini, Jansson & Elton (2007).

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 mandible (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. Bespoke software written in Visual Basic (Jones, University of Calgary) was used in this respect. The three landmarks used for matching the cranium and mandible configurations were eventually discarded and only the 86 anatomical landmarks used in the analyses.

Measurement error and estimates of a small number of missing landmarks (< 2% of all landmarks in 6.6% of specimens) were described in the Appendix and shown to have negligible effects on the analysis.

Functional and developmental modules

Analyses were performed first using all landmarks and then using only those on particular FDMs. Every step of the analysis was repeated for each FDM. All analyses were performed with separate sexes due to the large degree of sexual dimorphism observed in guenons (Cardini & Elton, in press a).

FDMs were defined following Hallgrímsson *et al*. (2004). These modules largely overlap with those of Cheverud (1995, 1996) and Ackermann & Cheverud (2000). Thus, four sets of FDMs (Fig. 2) were identified using the following criteria:

1. All available information. This was the most inclusive dataset, which employed all 86 landmarks to describe the form of the skull, with the mandible articulated on the occlusal plane.
2. Structural identity (i.e. cranium versus mandible). FDMs were here simply defined by being completely separate sets of bones.
3. Mode of ossification (Sperber, 2001) (i.e. chondrocranium versus dermatocranium). Much of the cranial base (most of the sphenoid, petrous portion of temporal, and the occipital bone) is formed by endochondral ossification. The remaining bones of the skull (face and cranial vault) all form by intramembranous ossification. (For brevity, the term 'cranial base' is used instead of the more appropriate but wordy 'chondrocranial portion of the cranial base' as a synonym of chondrocranium.)
4. Epigenetic effects on soft tissues and teeth (a). The dermatocranium can be divided into two main regions, the face and the cranial vault. The growth of the mammalian cranium is strongly influenced by the growth of the associated tissues, which, in turn, follow different patterns (Cheverud, 1995). The face is subjected to the potentially integrative action of the stresses generated by mastication and continues to grow until adulthood under the influence of general somatic factors (Cheverud, 1995; Hallgrímsson *et al*., 2004). The cranial vault is mostly influenced by the growth of the brain and completes its growth relatively early in postnatal life (Cheverud, 1995; Hallgrímsson *et al*., 2004). Thus, some lack of integration is expected between these two portions of the dermatocranium.
5. Epigenetic effects of soft tissues and teeth (b). Regions of the face can be further subdivided based on the direct influence of either teeth or masticatory muscles. The muscular portion of the cranium consists of the zygomatic arch and temporal fossa, which are hypothesized to be mostly influenced by the attachments of the main masticatory muscles, the temporalis and masseter (Cheverud, 1995; Hallgrímsson *et al*., 2004). The oral portion is composed of the regions of the skull surrounding the oral cavity and supporting the teeth, which are believed to be largely affected by the stress produced by mastication (Cheverud, 1995; Hallgrímsson *et al*., 2004). In the mandible, the oral and muscular regions, respectively, correspond to the horizontal and vertical ramus. The former supports the teeth and develops only in their presence (Cheverud, 1996; Lieberman, Pilbeam & Wood, 1996). The latter is the main area of insertion of muscles (temporalis on the coronoid process, masseter and medial pterygoid on the angular process, lateral pterygoid on the condyle), the action of which contributes to the development and maintenance of this region. The importance of muscle function and tooth growth for correct development of the muscular and oral regions of the skull are indicated by the severe bone anomalies in muscular dysgenesis mouse mutants (Herring, 1993) and the lack of alveolar process growth in the absence of teeth (Santana, Alvarez & Alabern, 1987; cited in Lieberman *et al*., 1996).

The landmarks included in each FDM are shown in Table 2. Four mandibular and five cranial landmarks were excluded from all analyses except that of the entire skull in order to reduce the weight of the numerous landmarks digitized on alveolar regions. The resulting configuration on teeth was virtually the same as in the analysis of baboon cranial variation by Frost *et al*. (2003).

We called the approach used for identifying FDMs a 'hierachical modular approach' because modules defined by different criteria correspond to a hierarchy from the most general and complex (the skull) to the most specific and localized (regions of the face and regions of the mandible).

Geometric morphometrics

Analyses were performed using geometric morphometrics (Bookstein, 1991; Dryden & Mardia, 1998; Zelditch *et al*., 2004) in the following computer programs: Morpheus (Slice, 1999), Morphologika, version 2.2 (O'Higgins & Jones, 2006), TPSSmall, version

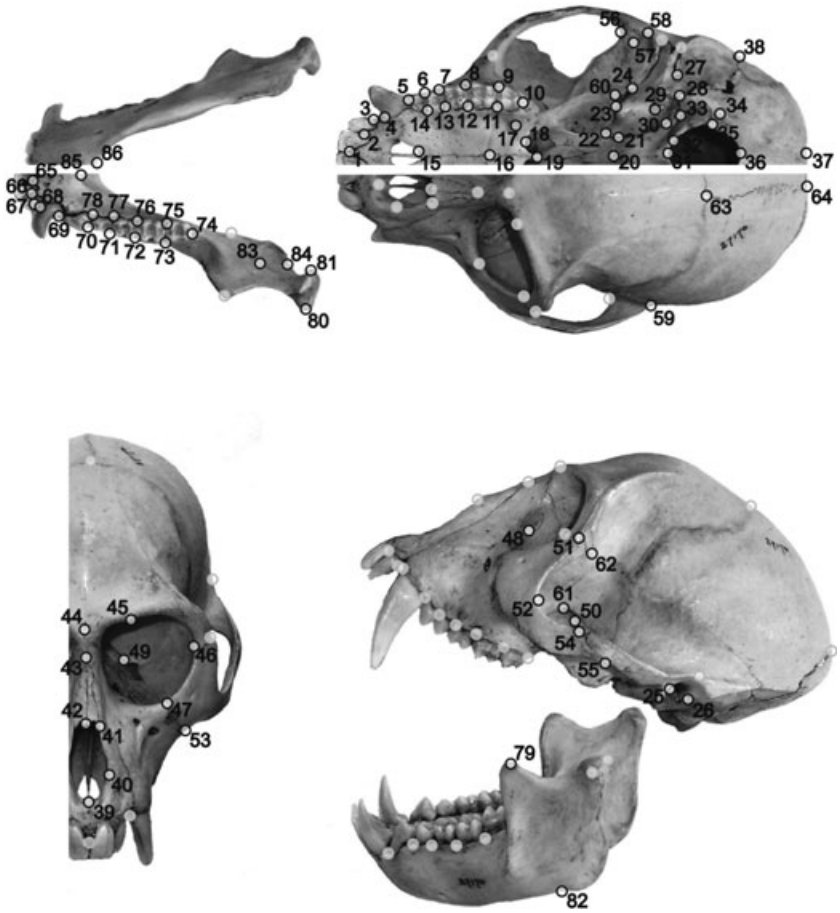

Figure 1. Landmark configuration including all landmarks. Definitions of anatomical landmarks are provided in Cardini *et al*. (2007).

Table 2. Functional and developmental module (FDM) landmark configurations (landmark numbers refer to Fig. 1)

Set	FDM	Number of landmarks	Included landmarks
1	Skull	86	All
2	Cranium	57	1–2, 4–5, 8, 10, 12, 15–64
2	Mandible	16	65–66, 68–69, 72, 74, 76, 79–86
3	Chondrocranium	18	20–24, 27–38, 60
3	Dermatocranium	42	1–2, 4–5, 8, 10, 12, 15–19, 24–26, 37–59, 61–64
4	Face	31	1–2, 4–5, 8, 10, 12, 15–19, 39–56, 59
4	Cranial vault	7	37–38, 44, 61–64
5	Oral	16	1–2, 4–5, 8, 10, 12, 15–19, 39–41, 50
5	Zygomatic	10	51–59, 61
5	Horizontal ramus	10	65–66, 68–70, 72, 74, 76, 85–86
5	Vertical ramus	6	79–84

1.20 (Rohlf, 2003), and NTSYSpc, version 2.2L (Rohlf, 2005). The form of an organism (or its organs) was 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 were removed, and size was standardized. This was achieved in the present study by optimally superim-

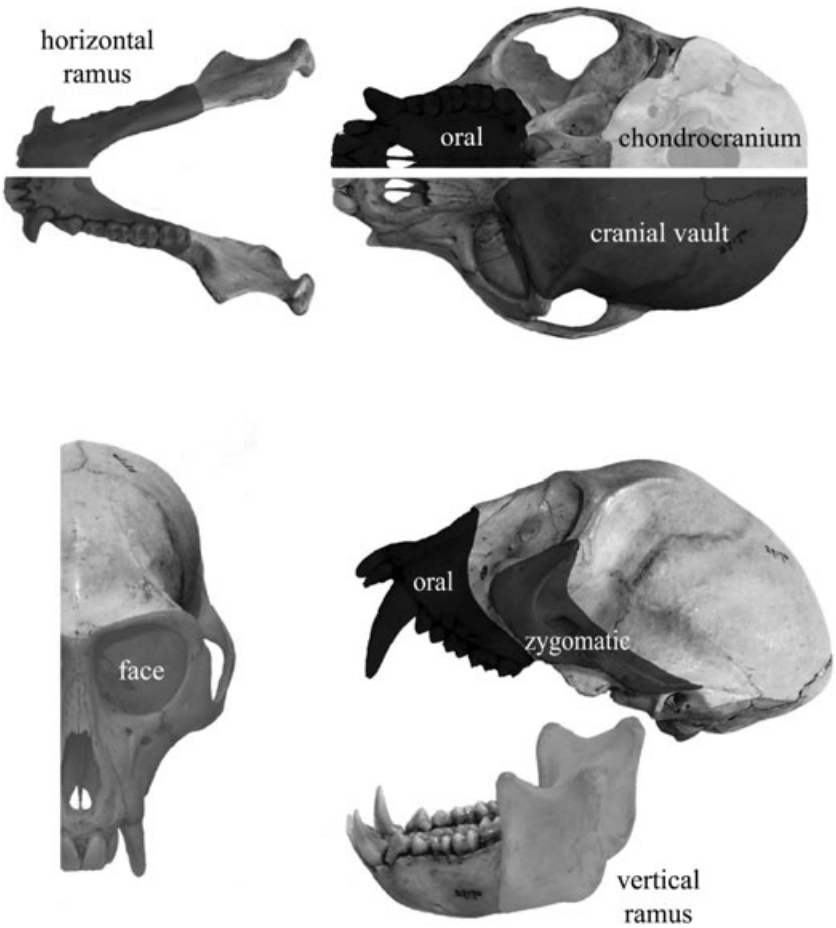

Figure 2. Functional and developmental modules of the skull, demarcated using different grey tones (dermatocranium not shown).

posing landmark configurations using generalized Procrustes analysis, which is based on a least-squares algorithm (Rohlf & 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, which is approximately the square root of the sum of squared distances between pairs of corresponding landmarks.

Statistical and phylogenetic analyses

Statistical analyses were performed using NTSYSpc (Rohlf, 2005). Genetic distances were reconstructed using the topology of the phylogenetic tree of Tosi *et al*. (2005: 63; fig. 3). Tree branches were proportional to the number of substitutions in the sequences of the approximately 9.3 kb fragment of X-chromosomal DNA used for inferring guenon phylogeny and could thus be used to build a matrix of pairwise genetic distances (COPH module of NTSYSpc, additive distances options). We assumed that genetic distances provided an accurate estimate of the true phylogenetic distances.

In the most recent taxonomic review of the Old World monkeys (Grubb *et al*., 2003), *Miopithecus* was divided into two species, *Miopithecus ogouensis* and *Miopithecus talapoin*. However, these two allopatric populations are very similar and were classified in the same species, *M. talapoin*, until recently (Kingdon, 1997). *Miopithecus ogouensis* was included in our analysis instead of *M. talapoin*, used by Tosi *et al*. (2005) for DNA sequencing. The error that this difference in taxonomic sampling might have introduced is negligible. This is because *Miopithecus* is the most distinctive guenon for skull morphology and the

two species are very similar (Cardini & Elton, in press b). Thus, using one or the other did not produce any appreciable difference. *Miopithecus ogouensis* was chosen because a larger sample of skulls was available in museum collections.

Specimens were Procrustes superimposed (one species at a time) and mean shapes computed for each species. The 15 species mean shapes that resulted were themselves superimposed and pairwise Procrustes distances computed. This was carried out for each FDM. To estimate the magnitude of the phylogenetic signal in a FDM, we computed the correlation between the matrix of mean shape Procrustes distances and the matrix of genetic distances. A Mantel test with 25 000 random permutations was used for testing the significance of a matrix correlation. This procedure is performed by randomly permuting rows and associated columns of one matrix, then calculating the correlation between the original and the permuted matrices. The observed correlation is significant if it is an outlier in the empirically derived distribution. The proportion of correlations larger than or equal to the observed correlation estimates is its P-value. A conservative significance threshold of 0.01 was chosen to avoid inflating the probability of type I errors in multiple tests.

The correlation between matrices of shape distances and the matrix of genetic distances can be influenced by the error in the estimate of species mean shapes (Cole *et al*., 2002; Cardini & Elton, 2007). To estimate this source of uncertainty, we used a bootstrap procedure (Cole *et al*., 2002; Caumul & Polly, 2006; Cardini & Elton, in press a). Thus, we randomly selected with replacement k individuals from each of the original species samples (where k is the number of specimens in the sample). Each bootstrapped sample was Procrustes superimposed and a new mean calculated. The resulting 15 bootstrap species mean shapes were themselves superimposed and the corresponding matrix of Procrustes distances computed. This was repeated 100 times, thus obtaining 100 Procrustes distance matrices of bootstrap mean shapes. Eventually, the correlation between each of the 100 Procrustes distance matrices and the matrix of genetic distances was computed. The dispersion of the distribution of correlation coefficients generated by the bootstrap procedure measured the precision (repeatability) of the observed coefficient. Precision was here defined as in Cole *et al*. (2002) as the correspondence between multiple sets of bootstrap samples taken from the same original sample. Thus, the SD of bootstrap correlations was calculated and used as an estimate of the SE of the observed matrix correlation.

The Mantel test compares the observed correlation against the null hypothesis that there is no relationship between two distance matrices. However, we may also want to know the probability of getting a matrix correlation as large as or larger than the one observed for a FDM described by q landmarks, if the same number of landmarks had been randomly selected from the original configuration of 86 landmarks. This tests the hypothesis that the phylogenetic signal is randomly distributed over the entire skull instead of being specific to a FDM. Thus, we randomly selected q landmarks from the original configuration of species mean shapes, performed a Procrustes superimposition, computed the matrix of Procrustes shape distances, and calculated its correlation with the genetic distances. This was repeated 1000 times to generate an empirical distribution of matrix correlations for random subsets of q landmarks. The observed correlation was significant if it was an outlier in the empirically derived distribution. This test was performed on FDMs with a significant matrix correlation in the Mantel test (i.e. those which putatively had a strong phylogenetic signal). As before, a conservative significance threshold of 0.01 was chosen. For brevity, we refer to this test as the test for matrix correlation in random subsets of landmarks.

Finally, the FDM with the strongest phylogenetic signal (largest significant matrix correlation) was used for reconstructing interspecific relationships. Phenetic relationships were summarized using a cluster analysis on the matrix of mean shape Procrustes distances. Different clustering algorithms were applied to the matrix of Procrustes shape distances (UPGMA, unweighted pair-group method using arithmetic average; UPGMC, unweighted pair-group method using centroid average; WPGMA, weighted pair-group method using arithmetic average; WPGMC, weighted pair-group method using centroid average). The goodness of fit of a cluster analysis was measured by the coefficient of cophenetic correlation. This was computed as the matrix correlation between the original distance matrix and the matrix of distances (cophenetic) based on the tree topology (Rohlf, 1970). Clustering algorithms were compared and the one with the highest coefficient of cophenetic correlation was chosen.

Phylogenetic relationships were estimated using two different algorithms. The first one was a maximum-likelihood (ML) method for quantitative traits that does not assume equal rates of change and treats shape variables as separate characters, optimizing the tree across them all (Felsenstein, 2004a). This method assumes a Brownian motion model of evolution, which is often appropriate for shape variables (Polly, 2003, 2008; Caumul & Polly, 2006) and uncorrelated characters. This second assumption was met by using the scores of the principal components of the mean shape variables standardized to a mean of zero

and variance equal to the proportion explained by the corresponding vector (Polly, 2008). Principal components were computed in NTSYSpc (Rohlf, 2005) and the CONTML module of PHYLIP, version 3.65 (Felsenstein, 2004b) was used for building the ML trees (options: input order of species randomized; ten global rearrangements). A second phylogenetic reconstruction was performed using a Neighbour-joining (NJ) algorithm in NTSYSpc (Rohlf, 2005). The NJ method was originally developed by Saitou & Nei (1987) for estimating phylogenetic trees from distance matrices by finding the pairs of taxa that minimize the total branch length at each stage of clustering (thus, having some relation to the criterion of minimum evolution; Felsenstein, 2004a). NJ is a clustering algorithm, like UPGMA and similar methods, but is aimed at phylogenetic reconstruction, allows outgroup rooting and has already been applied to matrices of Procrustes shape distances in other studies (Polly, 2001; Lockwood *et al.*, 2004). As for the phenograms produced by cluster analyses, the goodness of fit of the two methods was measured using their coefficients of cophenetic correlation (Polly, 2001). Phylogenetic trees were outgroup-rooted using *Allenopithecus nigroviridis*, traditionally considered a basal lineage in the cercopithecine radiation (Disotell, 2000).

After having selected the tree building algorithms (phenetic and phylogenetic) with the highest correlation to the original matrix of shape distances (see above), the repeatability of the resulting tree topologies was estimated using bootstraps. Thus, bootstrap mean shapes were computed as explained above, and the matrices of Procrustes shape distances of bootstrap replicates used for building trees. The percentage of trees in which each observed node grouping appeared was reported. Nodes with low percentages were strongly affected by sampling error (Cole *et al.*, 2002; Caumul & Polly, 2006; Cardini & Elton, in press b).

The same procedures described above for estimating the phylogenetic signal in shape were applied to centroid size. This was peformed because, even if size is generally considered to be more labile than shape, and thus unlikely to provide a strong phylogenetic signal, the potential for size to recover information on phylogenetic relationships cannot be excluded a priori.

RESULTS

Significance of the phylogenetic signal of FDMs

Matrix correlations between mean shape Procrustes distances and genetic distances are shown in Table 3, together with the corresponding SE and *P*-values. The SE measures the precision of the observed matrix correlation. The *P*-value is the Mantel test of significance. Observed matrix correlations were plotted as a column histogram in Figure 3.

The results for females and males were generally in good agreement. Matrix correlations within each set of FDMs can be summarized as:

1. All available information (skull shape): not significant;
2. Structural identity: mandible significant ($P < 0.01$);
3. Mode of ossification: chondrocranium significant ($P < 0.01$);
4. Epigenetic effects of soft tissues and teeth (a): not significant (cranial vault significant to $P < 0.01$ only in males);
5. Epigenetic effects of soft tissues and teeth (b): oral region of the cranium and vertical ramus of the mandible both significant to $P < 0.01$ (zygomatic-temporal fossa significant to $P < 0.01$ only in males).

Overall, the chondrocranium had the highest matrix correlation regardless of sex ($r > 0.6$). The second highest matrix correlation was in different FDMs for females (mandible) and males (oral region of the cranium), each being 0.6–1.5 points lower than the chondrocranium. The SE varied across FDMs and tended to be larger in FDMs that described more local features. This was suggested by large and negative bivariate correlations between SE and average FDM centroid size ($r_{\text{females}} = -0.825$; $r_{\text{males}} = -0.895$).

Mantel tests of significance were performed also for the correlations between the matrix of interspecific differences in mean FDM size and the matrix of genetic distances (Table 3). The phylogenetic signal in size was very weak in all FDMs and never reached statistical significance. Thus, size was not considered in further analyses.

The results of the test for matrix correlation in random subsets of landmarks are shown in Table 4. This was a test for the probability that observed FDM matrix correlations, which were significant in the Mantel test, were no larger than expected by chance if the same number of landmarks had been randomly selected from the configuration of all skull landmarks. The chondrocranium had significant ($P < 0.001$) matrix correlations in both sexes. The oral region of the cranium and the mandible were also significant ($P < 0.01$), but only in females. Thus, only the chondrocranium was used to reconstruct interspecific relationships of guenons.

Reconstructions of interspecific relationships based on FDMs

Among tree building algorithms (Table 5), UPGMA performed best in both sexes. However, WPGMA did

Table 3. Matrix correlations with bootstrap standard error (SE) and Mantel test of significance: genetic distances versus mean shape Procrustes distances and mean size differences

			Shape			Size		
Sex	Set	FDM	*r*	SE	*P*	*r*	SE	*P*
Female	1	Skull	0.427	0.018	0.0144	0.256	0.009	0.1357
	2	Cranium	0.428	0.018	0.0172	0.257	0.009	0.1307
	2	Mandible	0.563	0.052	0.0004*	0.260	0.008	0.1438
	3	Chondrocranium	0.630	0.049	0.0006*	0.256	0.023	0.1296
	3	Dermatocranium	0.426	0.019	0.0188	0.261	0.010	0.1202
	4	Face	0.442	0.023	0.0107	0.258	0.009	0.1289
	4	Cranial vault	0.418	0.042	0.0250	0.331	0.033	0.0660
	5	Oral	0.508	0.038	0.0020*	0.260	0.008	0.1261
	5	Zygomatic	0.443	0.033	0.0192	0.249	0.012	0. 1776
	5	Horizontal ramus	0.407	0.055	0.0132	0.286	0.016	0.1084
	5	Vertical ramus	0.496	0.061	0.0061*	0.211	0.017	0.2134
Male	1	Skull	0.421	0.014	0.0140	0.287	0.009	0.0846
	2	Cranium	0.419	0.015	0.0162	0.285	0.009	0.0843
	2	Mandible	0.472	0.036	0.0076*	0.280	0.009	0.0862
	3	Chondrocranium	0.682	0.029	0.0001*	0.346	0.018	0.0570
	3	Dermatocranium	0.414	0.015	0.0173	0.287	0.010	0.0810
	4	Face	0.408	0.019	0.0138	0.283	0.009	0.0844
	4	Cranial vault	0.484	0.030	0.0081*	0.304	0.018	0.0802
	5	Oral	0.528	0.033	0.0018*	0.290	0.011	0.0831
	5	Zygomatic	0.471	0.033	0.0082*	0.315	0.016	0.0680
	5	Horizontal ramus	0.264	0.045	0.0594	0.285	0.014	0.0876
	5	Vertical ramus	0.518	0.040	0.0048*	0.291	0.020	0.1040

FDM, functional and developmental module.
*$P < 0.01$.

Table 4. Test for matrix correlation in random subsets of landmarks ($P < 0.01$ in italics)

			Females		Males	
Set	FDM[a]	Number of landmarks	*r*	*P*	*r*	*P*
2	Mandible	16	0.563	0.001*	0.472	0.120
3	Chondrocranium	18	0.630	0.001*	0.682	0.001*
4	Cranial vault	7	0.418	–	0.484	0.084
5	Oral	16	0.508	0.004*	0.528	0.031
5	Zygomatic	10	0.443	–	0.471	0.195
5	Vertical ramus	6	0.496	0.062	0.518	0.039

[a]Test performed only on functional and developmental modules (FDMs) significant in the Mantel test (Table 3).
*$P < 0.01$.

almost equally well. Unsurprisingly, UPGMA and WPGMA tree topologies were virtually identical (data not shown). UPGMA phenograms for females and males are shown in Figures 4A and 5A, together with bootstrap proportions of node repeatability. The two main clades of the molecular phylogeny of Tosi *et al*. (2005) were correctly identified (Figs 4D, 5D). Thus, *Cercopithecus aethiops*, *Cercopithecus lhoesti*, and *Erythrocebus patas*, the terrestrial guenons, were discriminated from their sister clade of arboreal species, which includes all the other guenons except *A. nigroviridis* and *Miopithecus*. Beside differences in terminal branches, the most evident discrepancies between the phenograms of the chondrocranium and the molecular cladogram were the position of a member of the arboreal clade of Tosi *et al*. (2005),

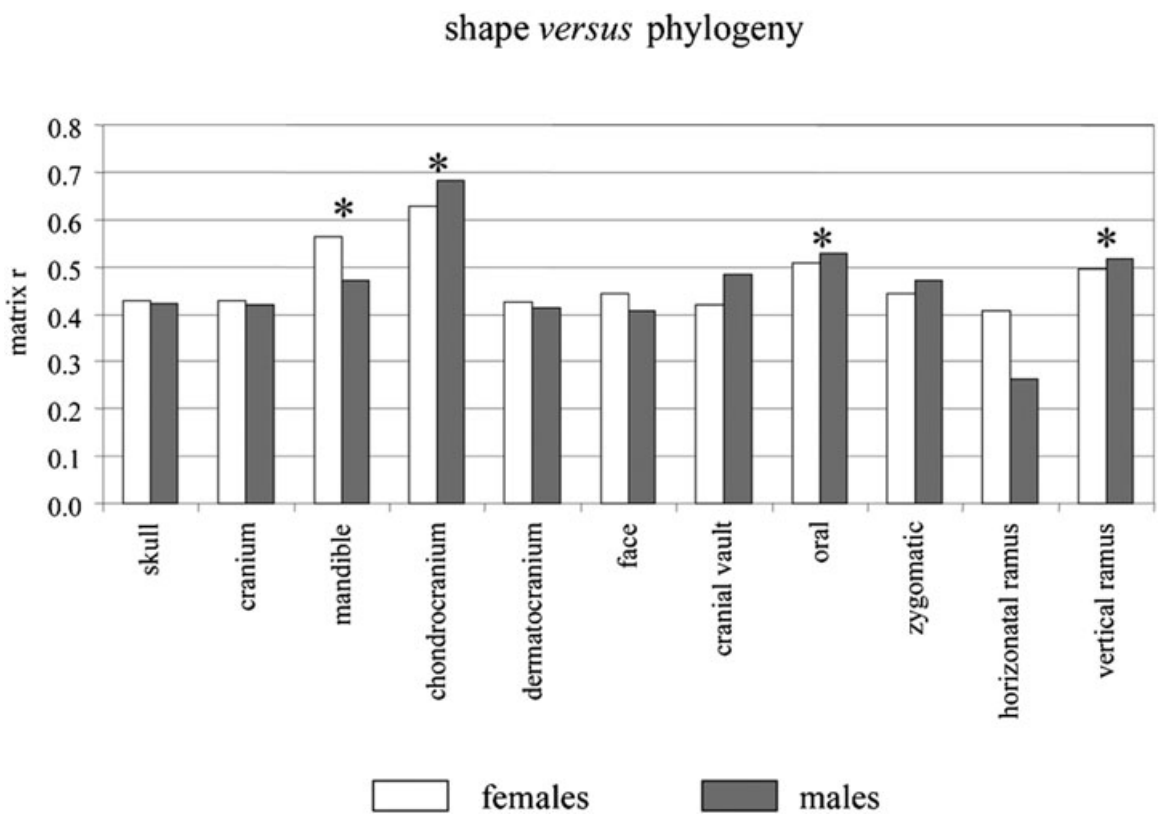

Figure 3. Column histogram of matrix correlations between genetic distances and Procrustes mean shape distances of all functional and developmental modules. Asterisks indicate significant Mantel test correlations in both sexes.

Table 5. Coefficients of cophenetic correlations of phenetic and phylogenetic trees

		r	
Type of reconstruction	Method	Females	Males
Phenetics	Complete	0.917	0.907
	Flexible	0.900	0.903
	Single	0.872	0.908
	UPGMA	0.924	0.926
	UPGMC	0.882	0.910
	WPGMA	0.920	0.923
	WPGMC	0.866	0.914
Phylogenetics	ML	0.906	0.949
	NJ	0.961	0.976

NJ, Neighbour-joining; ML, maximum likelihood; UPGMA, unweighted pair-group method using arithmetic average; UPGMC, unweighted pair-group method using centroid average; WPGMA, weighted pair-group method using arithmetic average; WPGMC, weighted pair-group method using centroid average.

Cercopithecus hamlyni, close to *A. nigroviridis*, and the proximity of the terrestrial guenons to the cluster *C. hamlyni–A. nigroviridis*. Also, *Miopithecus* did not show strong similarities to any other guenon. *Miopithecus* was near the root of the tree in the molecular phylogeny of Tosi *et al*. (2005). Repeatability was generally low with bootstrap proportions of node repeatability being on average 65.2% for females and 52.0% for males. However, clusters of terrestrial and arboreal species (*C. hamlyni* excluded) as well as the isolated *Miopithecus* were generally found in the majority of bootstrap replicate trees (repeatability $\geq$ 86.0%, except for male terrestrial guenons).

Surface rendering was used to visualize the main shape features of the chondrocranium of terrestrial and arboreal species, as well as those of the highly distinctive *Miopithecus* (Figs 4B, 5B). One of the most evident differences was the relatively narrow cranial base of terrestrial guenons compared to the enlarged cranial base of *Miopithecus*. The apparently marked reduction in the proportion of the supraoccipital region in arboreal species actually reflects its more vertical orientation.

Phylogenetic relationships were reconstructed using ML and NJ algorithms. The goodness of fit of

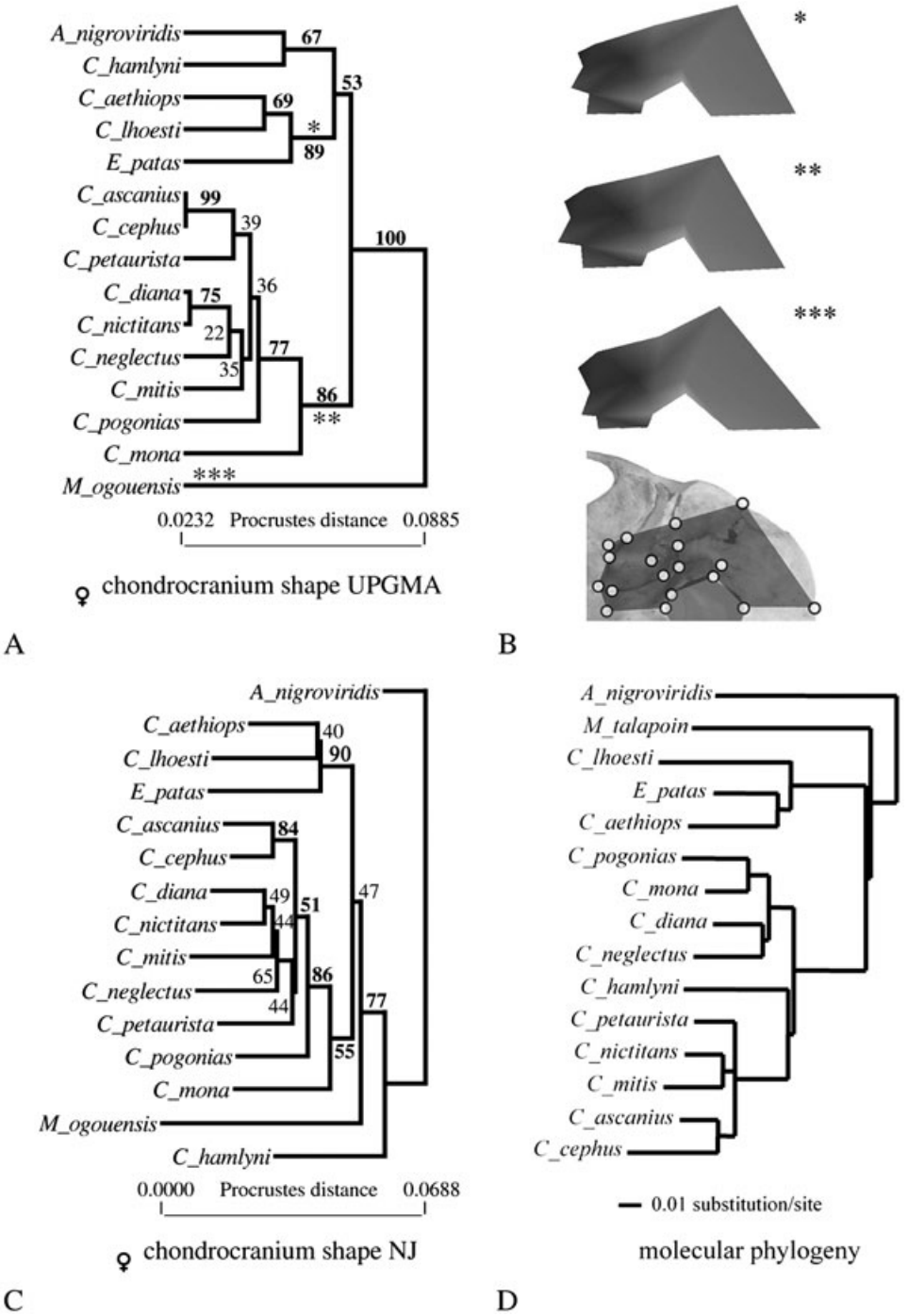

Figure 4. Guenon phenetic and phylogenetic relationships based on the shape of the female chondrocranium. A, unweighted pair-group method using arithmetic average (UPGMA) phenogram. B, surface rendering of the ventral view of the chondrocranium (variation magnified × 2) for the terrestrial (*) and arboreal (**) clusters, and the highly distinctive *Miopithecus* (***); the chondrocranium landmark configuration is also shown at the bottom. C, Neighbour-joining (NJ) phylogenetic tree. D, molecular phylogenetic tree (Tosi *et al*., 2005; redrawn). Bootstrap proportions of node repeatability (shown in bold if >50%) are shown on UPGMA and NJ trees.

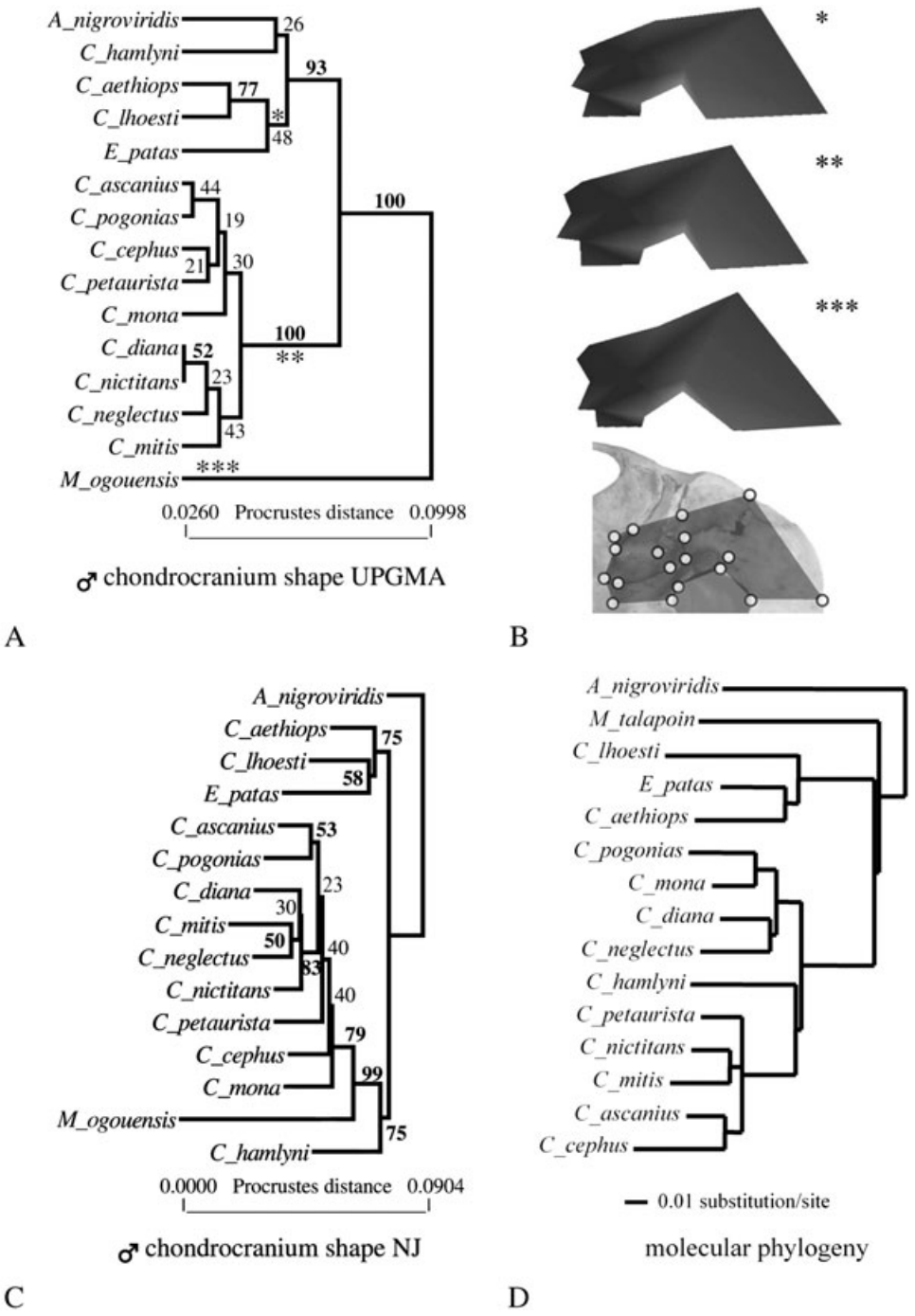

Figure 5. Guenon phenetic and phylogenetic relationships based on the shape of the male chondrocranium. Terrestrial (*) and arboreal (**) clusters, and the highly distinctive *Miopithecus* (***). NJ, Neighbour-joining. UPGMA, unweighted pair-group method using arithmetic average.

the phylogenetic reconstructions was compared using coefficients of cophenetic correlation (Table 5). The NJ algorithm performed better than the ML in both sexes. NJ cladograms for females and males are shown in Figures 4C and 5C, together with bootstrap proportions of node repeatability. As in the phenograms, clades of arboreal and terrestrial guenons were recognized well in the chondrocranium analysis. Arboreal and terrestrial clades were also correctly identified as sister groups in the female tree. *Cercopithecus hamlyni* was again outside the clade of arboreal guenons and close to the root of the tree.

Miopithecus females branched before the split of terrestrial and arboreal species, in good agreement with the molecular data, whereas males branched before the radiation of arboreal guenons. Compared to internodal distances, terminal branches were relatively long in both the terrestrial and arboreal clades, which is a pattern that bears some similarities to the molecular tree. As in the phenograms, bootstrap proportions of node repeatability were low (approximately 60% on average in both sexes), but consistently large (≥ 75%; Figs 4C, 5C) in the terrestrial and arboreal clades (with the exclusion of *C. hamlyni* and, possibly, *Cercopithecus mona*). In addition, a basal position of *Miopithecus* was found in most bootstrap replicate trees (repeatability ≥ 77%).

DISCUSSION

The major question addressed in the present study was whether any functional or developmental region of the guenon skull contained a significant phylogenetic signal. This was examined through combining geometric morphometrics and phylogenetics, and applying the 'hierarchical modular approach' previously used in evo-devo studies of morphological integration. Modularity is a central concept in evolutionary developmental biology. It can be seen as a manifestation of integration whenever integration within parts of a structure is stronger than among them. Wagner (1996) pointed out that modules may originate as semiautonomous parts of an organism, as distinct components of genetic networks, or as a set of traits covarying because of genetic pleiotropy and developmental interactions. Hallgrímsson *et al*. (2007) argued that integration in the skull is produced by a limited number of key developmental processes but also warned that '. . . patterns of phenotypic integration have a complex and sometimes indecipherable relationship to the underlying variational modularity that generated them'. Thus, they suggested that additional information, which can be obtained by studying the consequences of known developmental perturbations, may be needed to test hypotheses about the developmental determinants of modularity.

Mesenchymal differentiation, cartilage and brain growth, and muscle–bone interactions may differentially affect cranial structures and generate covariance. As some, but not all, of the most important covariance generating processes in the mammal skull, they represent the main criteria used in our study to define FDMs. Other processes that may have an important role in this context (e.g. neural crest migration and proliferation) are listed and briefly reviewed by Hallgrímsson *et al*. (2007). Neither our set of criteria for defining a priori FDMs, nor those in the list of Hallgrímsson *et al*. (2007) are exhaustive and tend to oversimplify the ontogeny of a complex structure like the skull. For example, if the embryologic origin of cells, from which cranial bones develop, had been included in our study as a criterion for defining FDMs, modules could have been defined that separate the frontal and parietal bones. This is because the former have been reported to be of neural crest origin (diencephalic/anterior mesencephalic) and the latter of mesodermal origin, with other bones such as the squamosal, alisphenoid, and pterygoid showing a mixed contribution from different rhombencephalic neural crest populations (Santagati & Rijli, 2003). Furthermore, as far as regions that originate from neural crest cells are concerned, uncertainties remain about whether the fate of those cells might reflect developmental instructions intrinsic to the neural crest and fixed before migration or neural crest cells might be instructed by surrounding tissues during or after migration; accumulating evidence indicates that the truth may lie somewhere between these two models (Pasqualetti & Rijli, 2002; Santagati & Rijli, 2003). Thus, the choice of criteria for defining modules and our imperfect knowledge of main processes involved in the development of the mammal skull may lead to define FDMs different from those used in our study and previous studies on modularity in primates (Cheverud, 1995, 1996; Ackermann & Cheverud, 2000; Hallgímsson *et al*., 2004).

FDMs that are clearly defined may be used to generate testable predictions (Lieberman, Ross & Ravosa, 2000), which can be modified later as comparative analyses and experimental studies to provide a better knowledge of the mechanisms which control the development of the skull. Thus, the way that we used a priori modules to test the strength of the phylogenetic signal is not meant to provide an exhaustive analysis of skull modularity in relation to phylogeny, but it is an example of how we can gain a better understanding of the phylogenetic signal in animal morphology by using a developmental framework to test hypotheses.

The chrondrocranium as a source of phylogenetic information

All analyses in the present study strongly suggested that a significant phylogenetic signal was present in chondrocranial shape. This FDM produced distance matrices that had the highest correlation with genetic distances, was significant in all tests, and contained enough phylogenetic information to discriminate the two major clades in the guenon radiation. No other FDM performed better, and the same pattern was consistently found in both females and males. That the chondrocranium performs well as an indicator of

phylogenetic relatedness was unexpected, as a previous analysis of guenon skulls showed that clusters of species had more to do with size groups (i.e. small, medium, large, and very large) than with clades (Cardini & Elton, in press a). This observation clearly raises the issue of whether there is a plausible biological explanation as to why a relatively simple region like the chondrocranium gives more information about phylogeny than the skull as a whole.

There is growing support in the literature for the concept that the cranial base as well as other cranial regions, such as the temporal (Lockwood *et al*., 2004), that include endochondral bones might reflect phylogeny better than do other parts of the skull. Lieberman *et al*. (2000) observed that the cranial base plays a key role in helping to spatially and functionally integrate different patterns of growth in various areas of the skull (brain, eyes, nasal cavity, oral cavity, pharynx), and revived the hypothesis (Olson, 1985; Shea, 1985, 1988; Lieberman *et al*., 2000) that the cranial base might be a more reliable indicator of phylogenetic relationships than are the face and cranial vault. Basicranial bones develop from cartilaginous precursors, so might be less sensitive to nongenetic factors (such as the mechanical forces of mastication) than are the bones of the cranial vault and face (Lieberman *et al*., 2000). In addition, as the cranial base is the first skull region to reach adult size, it is probably the area that is least affected by the growth of other parts of the skull (Lieberman *et al*., 2000). One consequence of these factors could be that, within the skull, the cranial base is most likely to have the strongest phylogenetic signal (Lieberman *et al*., 2000). However, previous analyses that compared inferences based on basicranial characters with those from the face or the cranial vault did not yield substantial differences (Lieberman *et al*., 1996). The importance of rigorously testing hypotheses on phylogenetic information of different sets of morphological characters using taxa with known molecular phylogenies has therefore been stressed (Lieberman *et al*., 2000). Our study of guenon skulls does precisely this.

At least one other primate study has highlighted the potential utility of parts of the skull that contain endochondral bones when reconstructing phylogeny. Lockwood *et al*. (2004) digitized three-dimensional landmarks on the hominoid temporal bone and, as in the present study, applied a NJ algorithm to the matrix of mean shape Procrustes distances to build a phylogenetic tree. When analyses with separate sexes were undertaken, similar patterns in females and males were found, and their trees corroborated the molecular consensus on African ape and human phylogeny by strongly supporting a *Pan-Homo* clade. This result is in striking contrast to the incongruencies found in previous studies of cranial characters (Collard & Wood, 2000, 2001). The success of the temporal bone analysis could be attributed to the accuracy of geometric morphometric data, the use of distance-based methods (as opposed to transformation of continuous characters into meristic ones), or the low level of homoplasy of the temporal bone because of the multitude of independent factors (including brain size and cognition, mastication, hearing, and posture) that might influence it (Lockwood *et al*., 2004). Further work is therefore required to explore whether the contrast between the results of Lockwood *et al*. (2004) and Collard & Wood (2001) is due to the utility of endochondral regions (because the latter did not focus on the cranial base alone), or whether it is because of methodological differences related to data collection, character coding, and tree building.

Retention and loss of phylogenetic signal in morphology: evolutionary scenarios

The present study strongly supported the hypothesis that the chondrocranium has the potential to be highly informative for reconstructing phylogenies, but several differences are evident when trees from chondrocranium shape are compared with the molecular cladogram. These can be interpreted in the context of various scenarios (Cole *et al*., 2002; Macholán, 2006) that explain the retention or loss of the phylogenetic signal in morphology. These include stochastic evolution, selection when lineages split to enter new adaptive zones, asymmetric adaptive shifts that result in convergence and homoplasy, 'star radiation', whereby multiple adaptive speciation occurs rapidly by the root of the tree, and speciation events within the same adaptive zone (Cole *et al*., 2002; Fig. 6). Two scenarios, 'star radiation' (Fig. 6D) and speciation within the same adaptive zone (Fig. 6E), do not appear to be compatible with any of the tree topologies from the chondrocranium, which consistently show at least one main dicotomy separating arboreal and terrestrial guenons. A further scenario, of stochastic evolution driven by genetic drift (Fig. 6A), is unlikely when the interspecific patterns recovered from the guenon chondrocranium are considered. Had divergence been a simple function of time subsequent to the last common ancestor, as implied by this scenario, we should have had a much stronger phylogenetic signal in the terminal branches of our trees, which were instead largely incongruent with the molecular phylogeny. By contrast, the scenario covering selection when lineages split to enter new adaptive zones (Fig. 6B) may be consistent with patterns for the arboreal and terrestrial clades. The phylogenetic signal in the chondrocranium discriminated these two major radiations,

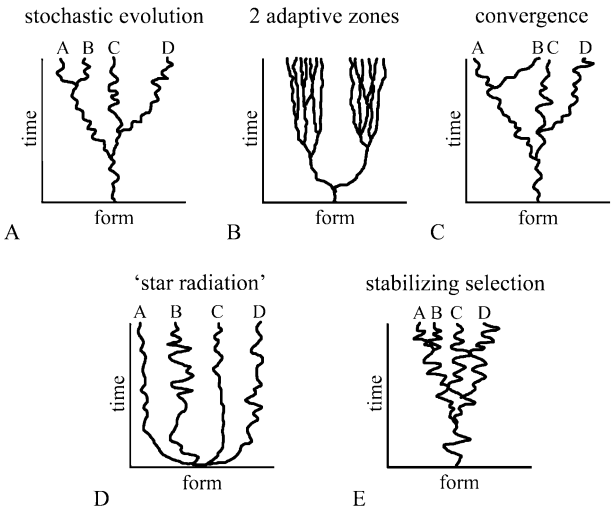

Figure 6. Scenarios of form evolution (modified from Cole *et al*., 2002). A, stochastic evolution with differences among lineages proportional to the time elapsed since their origin. B, colonization of new adaptive regions with further evolution occurring within the boundaries of each region. C, asymmetric adaptive shift with taxon B evolving similar form to taxon C due to convergence. D, 'star radiation' with rapid and multiple speciation events at the origin of the clade. E, speciation within the boundaries of a single adaptive zone (stabilizing selection; a similar pattern can be used to describe a scenario of evolutionary lability but fluctuations of form will then span a larger range of the morphospace).

but did not work within clades. This may be due to lineages crossing evolutionary paths within the constrained boundaries of their (terrestrial or arboreal) adaptive morphospace. Polly (2008) suggested that adaptive zones may act as a loosely constrained Ornstein–Uhlenbeck process, where stabilizing selection acts on phenotypes as if they were balls bound to an elastic band (Felsenstein, 1988). Thus, within populations, shapes wander back and forth on an adaptive peak with natural selection pulling them back towards the local optimum. The phylogenetic signal is likely to be erased by these evolutionary vagaries within a narrow region of morphospace (Polly, 2004, 2008). Morphological convergence (Fig. 6C) does not appear to be very strong in the chondrocranium. However, it is likely to have occurred in the guenon skull as a whole, where major shape clusters largely mirror interspecific differences in size (Cardini & Elton, in press b). Having shape similarities in skulls of similar size suggests a role for allometry in the evolution of guenon morphology. The chondrocranium, with its precocious development, is less affected by the growth of other parts of the skull (Lieberman *et al*., 2000) and thus potentially less strongly influenced by allometric changes.

Ecological pressures can act to significantly change morphology. There are numerous examples of this, but one of the most striking is seen in the pinnipeds. Part of the carnivore clade, the pinnipeds evolved very distinctive shapes, which are difficult to associate with those of their closest relatives, by escaping the boundaries of their ancestral terrestrial adaptive zone (Polly, 2008). A less dramatic, but nonetheless analogous, process might have occurred in guenons. In our guenon sample, the shape of the chondrocranium of *Miopithecus* was highly distinctive. This strong autoapomorphism may be related to its ancient evolutionary origin, some 8.6 Mya, according to the molecular clock of Tosi *et al*. (2005). It is, however, not unlikely that the 'aberrant' (Verheyen, 1962) shape of *Miopithecus* may have arisen as a by-product of a strong reduction in size driven by ecological divergence. The ecological conditions under which selection for small body size occurred in *Miopithecus* are still largely unexplored. Dietary adaptation might be one explanation, given the observation that over one-third of the talapoin diet comprises insects (Gautier-Hion, 1978), and the expectation that insectivorous primates are considerably smaller than those that are frugivorous or folivorous.

Cercopithecus hamlyni, which probably diverged around 4.6 Mya (Tosi *et al*., 2005), is another guenon species that showed a marked departure from expectations of shape differences based on molecular data. This species was clearly different to other arboreal guenons but bore some resemblance to *A. nigroviridis*. Morphological convergence through similarities in ecology might explain this but, at present, this cannot be ascertained because too little is known about the ecology of *C. hamlyni*, whose present populations are mostly restricted to a small number of mountain forests in eastern Congo.

Reproducibility of results

One critical issue in reconstructing phylogeny is whether it is reproducible. We addressed this issue by bootstrapping species samples to estimate the error in mean shapes and its effect on matrix correlations and tree topologies. Large errors can be expected when within-species variation is high relative to between-species variation, and also when sample sizes are small (Caumul & Polly, 2006). By using repeated randomized selection experiments in a large sample of *C. aethiops*, we have shown that the error in the mean shape for samples of ten individuals can be, on average, as large as 37% of the interspecific distance between the mean shapes of *C. aethiops* and *Cercopithecus mitis* (Cardini & Elton, 2007). In the present study, our sample size was ≥ 34 on average, which is larger than in most

taxonomic studies on mammals. However, even with this reasonable sample, the sampling error accounted for appoximately 7–8% ($100 \times SE/r_{matrix}$) of the observed correlations between matrices of shape distances and the genetic distance matrix. Bootstrap proportions of node repeatability were lower than 70% in 50–70% of tree nodes. These findings were consistent with observations on marmots (Caumul & Polly, 2006), as well as in our previous work on guenons (Cardini & Elton, in press b), and strongly suggest that morphological analyses of closely-related species should always provide an estimate of the effect of sampling error.

The estimation of sampling error produced a second interesting outcome. On the whole, configurations of less than 20 landmarks (such as those describing local aspects of the skull like the chondrocranium or oral regions) had SE of matrix correlations two-fold greater than those observed in the regions with configurations exceeding 30 landmarks (e.g. the face and dermatocranium). Measurement error tends to be relatively larger in smaller structures (Polly, 1998) and, as strongly suggested by the large negative correlations between SE and FDM size in the present study, this also affects mean shapes. Thus, the chondrocranium had the strongest phylogenetic signal but it was also one of the FDMs with the lowest reproducibility. If this is a general finding, the potential of localized anatomical regions to provide 'good' phylogenetic characters is unlikely to be realized unless errors in means can be controlled by reducing measurement error and increasing sample size.

The reproducibility of results can also depend on the choice of landmarks used to capture the shape of a FDM. Lockwood *et al*. (2004) and Couette, Escarguel & Montuire (2005) addressed this issue by bootstrapping shape variables. This is the most widely used test for the effect of character sampling on tree topologies in cladistic analyses of meristic characters. However, Caumul & Polly (2006: 2464) observed that, when estimating the support for the nodes in a tree using 'ordinary phylogenetic data, the characters are sampled with replacement to generate a new dataset [but] for morphometric data, the characters are scores whose values depend on the ordination of the mean shapes'. Without discussing if and to what extent bootstrapping shape variables might be (in)appropriate, we suggest an alternative procedure, also borrowed from the cladistic literature (Felsenstein, 2004a) and easily applicable to geometric morphometric data. A jack-knife procedure can be applied to the matrix of landmark coordinates in order to produce replicate landmark configurations where different 'bits' of information are excluded from the analysis. Thus, one at a time, each landmark (more precisely its x, y, and z coordinates) is excluded, the species mean shapes are Procrustes superimposed, and a matrix of pairwise Procrustes distances is computed. This is repeated as many times as the total number of landmarks in the original configuration. The corresponding matrices of variables or distances are then used for tree building. The proportions of clusters found in the jack-knife replicate trees are the estimates of the 'character' (i.e. landmark configuration) support of tree nodes. Had this been applied to the guenon chondrocranium, the 'character' support would have been, on average, approximately 20% higher than our bootstrap proportions of node repeatability. Random replacement of shape coordinates, as in the bootstrap procedure used in other studies, is very likely to produce results similar to our jack-knife procedure. Indeed, as was the case in our jack-knife example, both Lockwood *et al*. (2004) and Couette *et al*. (2005) found very high percentages of tree node support. Although we do not argue against the utility of estimating uncertainties in tree nodes due to 'character sampling' (which we would do, however, using a jack-knife when applied to shape coordinates), we do suggest that the effects of sampling error should also be estimated. Since this could be large (Cole *et al*., 2002; Caumul & Polly, 2006; Cardini & Elton, 2007), its consequences could potentially be more profound than those due to uncertainties in 'character sampling'.

How much information to use?

Another pressing issue when reconstructing evolutionary relationships using phylogenetic analysis is determining the 'optimal' amount of information to include. Polly (2001) argued by analogy with molecular data that homoplasy was less likely to confound the phylogenetic signal in morphological features that contained more information by virtue of complexity. This view is reminiscent of a point made by Felsenstein (1988) in his seminal study of phylogeny and quantitative characters. Although Felsenstein discusses various problems that are encountered when using morphometric data for inferring phylogenies, he also states that the phylogeny will converge on the correct one if enough characters are accumulated whose correlation is, on average, low (Felsenstein, 1988: 460). Finding characters that are are mostly uncorrelated is not an easy task but it does appear at times that increasing the amount of information available for phylogenetic inference may indeed be beneficial. In a summary of results from several studies of primate bone morphology, Oxnard (2000) observed that morphometric comparisons of individual skeletal units (such as arms, limbs or teeth) tend to produce

clusters that indicate functional convergences of anatomical parts. By contrast, when variables from different anatomical regions are combined in a single analysis, separations of species mostly reflect evolutionary relatedness. Oxnard (2000) explained these results using a simple model whereby functional and phylogenetic components of morphology (and also their interactions) are described in terms of units of information. Phylogenetic information within a structure is relatively small and its signal tends to be overcome by function. However, when several structures are analysed together, the functional units are unlikely to provide a signal in a consistent direction because functions tend to be different. By contrast, as the phylogenetic units all give a similar indication of evolutionary relatedness, their contribution becomes proportionally larger, and phylogenetic groups emerge (Oxnard, 2000).

Other studies (Lieberman *et al*., 1996, 2000; Lockwood *et al*., 2004) have not echoed the confidence shown by some workers in adding more information to increase the strength of the phylogenetic signal from morphology. The results obtained by Caumul & Polly (2006) did not indicate that more complexity provides better phylogenies; in their study, the simple outline of molar teeth performed better than the mandible. It has also been argued that anatomical regions that are thought to be more heritable and less influenced by epigenetic responses to external stimuli should be targeted, and their phylogenetic signal tested against known molecular phylogenies (Lieberman *et al*., 1996, 2000; Lockwood *et al*., 2004), with Lockwood *et al*. (2004) finding a strong phylogenetic signal in a relatively small region of the skull. Traits with high heritability are also highly susceptible to selection and therefore may be more likely to show convergence or parallelism. However, it is possible that traits showing high heritability and small changes in relation to external stimuli are not highly plastic and therefore their shape does not depend strongly on individual life events.

The main outcome of the present study (i.e. that the chondrocranium performs particularly well in phylogenetic reconstruction) is largely in agreement with the notion that 'bigger' (specifically, more complex) is not necessarily 'better'. None of the FDMs with the largest amount of information (skull, cranium, dermatocranium, face) had a phylogenetic signal stronger than the chondrocranium. Thus, dissecting complex anatomies into simpler units may sometimes help phylogenetic reconstructions. This notwithstanding, the issue of reproducibility discussed above must be seriously considered when focusing on a smaller region like the chondrocranium. Doing so might increase the chances of detecting a phylogenetic signal, but it is also likely to reduce the reproducibility of results. In the present study, FDMs which describe local features not only had larger SE of matrix correlations, but also smaller bootstrap proportions of node repeatability (data not shown). Similarly, Caumul & Polly (2006) obtained average bootstrap proportions of 56.5%, 52.8%, and 39.0% for the cranium, mandible, and molars, respectively. Thus, a similar trend of larger errors in smaller/less complex structures was found in studies of monkeys and marmots. If this is a general finding, its direct implication is again that the applicability of the use of FDMs such as the chondrocranium for phylogenetic inference in closely-related taxa may be limited by measurement and sampling errors.

Conclusions

Four main conclusions can be drawn from our study. The first is that the chondrocranium might be less susceptible to epigenetic factors and thus more informative about phylogeny than are other parts of the skull, at least in guenons. Whether this holds for other primates or orders of placental mammals needs to be investigated. Among primates, papionins offer an excellent model for further tests, as they are the sister group of guenons and have a well supported molecular phylogeny at the genus level. That previous morphological studies failed to detect a strong phylogenetic signal (Collard & Wood, 2000, 2001) serves to increase their utility for such a test, as further research might help to determine whether the choice of method used to collect data and subsequent character coding impacts significantly upon the success of the phylogenetic reconstruction. Second, although it has been suggested that complex morphologies have greater potential to reveal phylogeny than simpler structures (Felsenstein, 1988; Polly, 2001), there is no evidence to support this in the current study of guenons. However, sampling error tends to be larger in smaller/less complex structures, a finding that is supported by analyses of species from different orders of mammals. If confirmed in other groups, the utility of less complex regions, such as specific FDMs, for phylogenetic inference may be hampered by the low repeatability of results. This links closely to a third finding, that sampling error strongly affects phylogenetic reconstructions. Since uncertainties in tree topologies due to sampling error can be larger than those due to character sampling, we argue that estimates of sampling error should always be provided in phylogenetic analyses of form. Finally, using results from the analysis of the chondrocranium and comparing tree topologies from morphological and molecular data in the context of different evolutionary sce-

narios, we suggest that the main morphological differences in guenons evolved at the time of the divergence between the terrestrial and arboreal clades. Further morphological evolution probably occurred within the narrow boundaries of the adaptive zone of each clade, where lineages cross their evolutionary paths and thus erase the phylogenetic signal among closely-related species. Our analysis therefore exemplifies how studying the potential of different skull modules as proxies for phylogenetic divergence in modern lineages may provide a framework for modelling the morphological evolution of present and fossil species.

ACKNOWLEDGEMENTS

We are greatly in debt to S. Cobb (University of Hull), K. Kupczik (Max-Planck Institute for Evolutionary Anthropology, Leipzig), J. M. Cheverud (Washington University School of Medicine, St Louis) and B. Hallgrímsson (University of Calgary) for their advice on aspects of primate skull anatomy and to P. O'Higgins (University of York) and P. D. Polly (Indiana University, Bloomington) for their comments on a earlier version of this manuscript. We also thank the Editor and two reviewers for their comments on this manuscript. Our most sincere gratitude is extended to all the museum curators and collection managers who allowed us to study their specimens and who greatly helped us during data collection, thus making this work possible. Several other people contributed directly or indirectly to our work. Most of them, hopefully all, are acknowledged in Cardini & Elton (in press b). To all of them, once more, we extend our deepest gratitude. This study was funded by a grant from The Leverhulme Trust.

REFERENCES

Ackermann RR, Cheverud JM. 2000. Phenotypic covariance structure in tamarins (genus *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. *American Journal of Physical Anthropology* **111:** 489–501.

Adams DC, Slice DE, Rohlf FJ. 2004. Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology* **71:** 5–16.

Bookstein FL. 1991. *Morphometric tools for landmark data*. Cambridge, MA: Cambridge University Press.

Cardini A. 2003. The geometry of marmot (Rodentia: Sciuridae) mandible: phylogeny and patterns of morphological evolution. *Systematic Biology* **52:** 186–205.

Cardini A, Elton S. 2007. Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology* **126:** 121–134.

Cardini A, Elton S. in press a. Variation in guenon skulls II: sexual dimorphism. *Journal of Human Evolution* (in press).

Cardini A, Elton S. in press b. Variation in guenon skulls I: species divergence, ecological and genetic differences. *Journal of Human Evolution* (in press).

Cardini A, Jansson A-U, Elton S. 2007. A geometric morphometric approach to the study of ecogeographic and clinal variation in vervet monkeys. *Journal of Biogeography* **34:** 1663–1678.

Cardini A, O'Higgins P. 2004. Patterns of morphological evolution in *Marmota* (Rodentia, Sciuridae): geometric morphometrics of the cranium in the context of marmot phylogeny, ecology, and conservation. *Biological Jorunal of the Linnean Society* **82:** 385–407.

Cardini A, Thorington RW Jr. 2006. Post-natal ontogeny of the marmot (Rodentia, Sciuridae) cranium: allometric trajectories and species divergence. *Journal of Mammology* **87:** 201–216.

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.

Cheverud JM. 1995. Morphological integration in the saddle-back tamarin (*Saguinus fuscicollis*) cranium. *American Naturalist* **145:** 63–89.

Cheverud JM. 1996. Developmental integration and the evolution of pleiotropy. *American Zoologist* **36:** 44–50.

Cole TM III, Lele S, Richtsmeier JT. 2002. A parametric bootstrap approach to the detection of phylogenetic signals in landmark data. In: Macleod N, Forey P, eds. *Morphology, shape and phylogeny*. New York, NY: Taylor & Francis, 194–219.

Collard M, Wood B. 2000. How reliable are human phylogenetic hypotheses? *Proceedings of the National Accademy of Sciences of the United States of America* **97:** 5003–5006.

Collard M, Wood B. 2001. Homoplasy and the early hominid masticatory system: inferences from analyses of extant hominoids and papionins. *Journal of Human Evolution* **41:** 167–194.

Couette S, Escarguel G, Montuire S. 2005. Constructing, bootstrapping, and comparing morphometric and phylogenetic trees: a case study of new world monkeys (platyrrhini, primates). *Journal of Mammology* **86:** 773–781.

Disotell TR. 2000. Molecular systematics of the Cercopithecidae. In: Whitehead PF, Jolly CJ, eds. *Old world monkeys*. Cambridge: Cambridge University Press, 29–56.

Dryden IL, Mardia KV. 1998. *Statistical shape analysis*. New York, NY: John Wiley and Sons.

Fadda C, Corti M. 2001. Three-dimensional geometric morphometrics of Arvicanthis: implications for systematics and taxonomy. *Journal of Zoology Systematics and Evolutionary Research* **39:** 235–245.

Felsenstein J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* **19:** 445–471.

Felsenstein J. 2004a. *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.

Felsenstein J. 2004b. *PHYLIP (phylogeny inference package)*, version 3.6. Distributed by the author. Seattle, WA: Department of Genome Sciences, University of Washington.

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. *Anatomical Record* **275A:** 1048–1072.

Gaubert P, Veron G. 2003. Exhaustive sample set among Viverridae reveals the sister-group of felids: the linsangs as a case of extreme morphological convergence within Feliformia. *Proceedings of the Royal Society of London Series B, Biological Sciences* **270:** 2523–2530.

Gautier J-P, Vercauteren Drubbel R, Deleporte P. 2002. Phylogeny of the *Cercopithecus lhoesti* group revisited: combining multiple characher sets. In: Glenn ME, Cords M, eds. *The guenons: diversity and adaptation in African monkeys*. New York, NY: Kluwer Academic/Plenum Publishers, 37–48.

Gautier-Hion A. 1978. Food niches and coexistance in sympatric primates in Gabon. In: Herbert DJ, ed. *Recent advances in primatology chivers*, Vol. II. New York, NY: Academic Press, 270–286.

Goswami A. 2006. Cranial modularity shifts during mammalian evolution. *American Naturalist* **168:** 2700–2280.

Grubb P, Butynski TM, Oates JF, Bearder SK, Disotell TR, Groves CP, Struhsaker TT. 2003. Assessment of the diversity of African primates. *International Journal of Primatology* **24:** 1301–1357.

Guy F, Brunet M, Schmittbuhl M, Viriot L. 2003. New approaches in hominoid taxonomy: morphometrics. *American Journal of Physical Anthropology* **121:** 198–218.

Halanych KM. 2004. The new view of animal phylogeny. *Annual Review of Ecology Evolution and Systematics* **35:** 229–256.

Hallgrímsson B, Lieberman DE, Young NM, Parsons T, Wat S. 2007. Evolution of covariance in the mammalian skull. In: *Tinkering: the microevolution of development*. Novartis Foundation Symposium 284. Chichester: Wiley.

Hallgrímsson B, Willmore K, Dorval C, Cooper DM. 2004. Craniofacial variability and modularity in macaques and mice. *Journal of Experimental Zoology Part B. Molecular and Developmental Evolution* **302:** 207–225.

Herring SW. 1993. Formation of the vertebrate face: epigenetic and functional influences. *Amererican Zoologist* **33:** 472–483.

Jensen RJ. 2003. The conundrum of morphometrics. *Taxon* **52:** 663–671.

Kingdon JS. 1997. *The kingdom field guide to African mammals*. London: Academic Press.

Koepfli K-P, Jenks SM, Eizirik E, Zahirpour T, Van Valkenburgh B, Wayne RK. 2006. Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Molecular Phylogenetics and Evolution* **38:** 603–620.

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biology* **4:** 537–544.

Lieberman DE, Pilbeam DR, Wood BA. 1996. Homoplasy and early *Homo*: an analysis of the evolutionary relationships of *H. habilis sensu stricto* and *H. rudolfensis*. *Journal of Human Evolution* **30:** 97–120.

Lieberman DE, Ross CF, Ravosa MJ. 2000. The primate cranial base: ontogeny, function, and integration. *Yearbook of Physical Anthropology* **43:** 117–169.

Lockwood CA, Kimbel WH, Lynch JM. 2004. Morphometrics and hominoid phylogeny: support for a chimpanzee-human clade and differentiation among great ape subspecies. *Proceedings of the National Accademy of Sciences of the United States of America* **101:** 4356–4360.

Macholán N. 2006. A geometric morphometric analysis of the shape of the first upper molar in mice of the genus *Mus* (Muridae, Rodentia). *Journal of Zoology* **270:** 672–681.

MacLeod N. 2002. Phylogenetic signals in morphometric data. In: MacLeod N, Forey P, eds. *Morphology, shape and phylogeny*. New York, NY: Taylor & Francis, 100–138.

Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de Jong Wilfried W, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409:** 610–614.

Milinkovitch MC, Thewissen JGM. 1997. Even-toed fingerprints on whale ancestry. *Nature* **388:** 622–624.

Milne N, O'Higgins P. 2002. Inter-specific variation in *Macropus* crania: form, function and phylogeny. *Journal of Zoology* **256:** 523–535.

Monteiro LR, dos Reis SF. 2005. Morphological evolution in the mandible of spiny rats, genus *Trinomys* (Rodentia: Echimyidae). *Journal of Zoology Systematics and Evolutionary Research* **43:** 332–338.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryderk OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* **409:** 614–618.

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). *Biological Journal of the Linnean Society* **80:** 385–396.

O'Higgins P, Jones N. 2006. *Morphologika 2.2. Tools for shape analysis*. York: Hull York Medical School, University of York. Available at http://www.york.ac.uk/res/fme/.

Olson TR. 1985. Cranial morphology and systematics of the Hadar hominids and *Australopithecus africanus*. In: Delson E, ed. *Ancestors: the hard evidence*. New York, NY: AR Liss, 102–119.

Oxnard CE. 2000. Morphometrics of the primate skeleton and the functional and developmental underpinnings of species diversity. In: O'Higgins P, Cohn MJ, eds. *Development, growth and evolution*. London: Academic Press, 235–263.

Pasqualetti M, Rijli FM. 2002. The plastic face. *Nature* **416:** 493–494.

Pilbeam D. 2000. Hominoid systematics: the soft evidence. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 10684–10686.

Polly PD. 1998. Variability in mammalian dentitions: size-related bias in the coefficient of variation. *Biological Journal of the Linnean Society* **64:** 83–99.

Polly PD. 2001. On morphological clocks and paleophylogeography. *Genetica* **112/113:** 339–357.

Polly PD. 2003. Paleophylogeography: the tempo of geographic differentiation in marmots (Marmota). *Journal of Mammology* **84:** 369–384.

Polly PD. 2004. On the simulation of the evolution of morphological shape: multivariate shape under selection and drift. *Palaeontologica Electronica* **7.2.7A**. Available at http://palaeo-electronica.org/paleo/2004p2/evo/issue2p04.htm.

Polly PD. 2008. Adaptive zones and the pinniped ankle: a 3D quantitative analysis of Carnivoran tarsal evolution. In: Sargis E, Dagosto M, eds. *Mammalian evolutionary morphology: a tribute to Frederick S Szalay*. Dordrecht: Springer, 1–40.

Rohlf FJ. 1970. Adaptive hierarchical clustering schemes. *Systematic Zoology* **19:** 58–82.

Rohlf FJ. 2003. *Tpssmall*. Stony Brook, NY: Department of Ecology and Evolution. State University of New York. Available at http://life.bio.sunysb.edu/morph/.

Rohlf FJ. 2005. *Ntsyspc*, Version 2.20L. Setauket, NY: Exeter Software.

Rohlf FJ, Marcus LF. 1993. A revolution in morphometrics. *Trends in Ecology and Evolution* **8:** 129–132.

Rohlf FJ, Slice DE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* **39:** 40–59.

Rüber L, Adams DC. 2001. Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. *Journal of Evolutionary Biology* **14:** 325–332.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for constructing phylogenetic trees. *Molecular Biology and Evolution* **4:** 406–425.

Santagati F, Rijli FM. 2003. Cranial neural crest and the building of the vertebrate head. *Nature Reviews Neuroscience* **4:** 806–818.

Santana SM, Alvarez FP, Alabern C. 1987. Agnathia and associated malformations. *Dysmorphology and Clinical Genetics* **1:** 58–72.

Shea BT. 1985. On aspects of skull form in African apes and orangutans, with implications for hominoid evolution. *American Journal of Physical Anthropology* **68:** 329–342.

Shea BT. 1988. Phylogeny and skull form in the hominoid primates. In: Schwartz JH, ed. *The biology of the orangutan*. New York, NY: Oxford University Press, 233–246.

Singleton AM. 2002. Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). *Journal of Human Evolution* **42:** 547–578.

Slice DE. 1999. *Morpheus* (beta version). Stony Brook, NY: Department of Ecology and Evolution, State University of New York.

Slice DE. 2005. Modern morphometrics. In: Slice DE, ed. *Modern morphometrics in physical anthropology*. Kluwer, NY: Sperber, 1–45.

Sperber GH. 2001. *Craniofacial development*. Hamilton, BC: Decker Inc.

Springer MS, Cleven GC, Madsen O, de Jong WW, Waddell VG, Amrine HM, Stanhope MJ. 1997. Endemic African mammals shake the phylogenetic tree. *Nature* **388:** 61–64.

Stanhope MJ, Waddell VG, Madsen O, De Jong W, Hedges SB, Cleveni GC, Kaoi D, Springer MS. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 9967–9972.

Tosi AJ, Detwiler KM, Disotell TR. 2005. X-chromosomal window into the evolutionary history of the guenons (Primates: Cercopithecini). *Molecular Phylogenetics and Evolution* **36:** 58–66.

Verheyen WN. 1962. Contribution a la craniologie comparee des primates. Les genres *Colobus* Illiger 1811 et *Cercopithecus* Linne 1758. *Musee Royal de l'Afrique Centrale, Tervuren, Annales* **8:** 1–255.

Wagner GP. 1996. Homologues, Natural Kinds and the Evolution of Modularity. *American Zoologist* **36:** 36–43.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004. *Geometric morphometrics for biologists: a primer*. London: Elsevier Academic Press.

APPENDIX

Measurement error

One specimen was measured 12 times, without changing its position, to estimate the landmark absolute precision. The SD of the Cartesian coordinates of a landmark was 0.8 mm on average. This included both human and instrument errors (the digitizer accuracy is 0.38 mm according to the producer) and it is quite small relative to the size of guenon skulls (maximum range of cranial length: 63.9–153.7 mm).

To estimate the overall precision of size and shape variables, six specimens were repeatedly measured over six consecutive days. Every day, the entire data collection process was repeated in order to include all sources of error (positioning, digitization, match of cranium, and mandible configurations). The sum of variances of the registered skull coordinates was computed for each specimen and then averaged (average error of a specimen). The ratio between the error shape variance and the shape variance in the analysis sample gives an approximate estimate of the relative amount of shape variation accounted for by measurement error. This ratio computed separately for each species and sex and expressed as a percentage was on average $\leq 8.0\%$. If a pooled sample of all species and both sexes was used, the ratio becomes 4.4%.

Precision of centroid size was estimated by computing the coefficient of variation ($100 \times$ SD/mean) of the six repetitions of each specimen. The coefficient was always $< 0.2\%$.

An estimate of the precision comparable to those of traditional caliper measurements was obtained by considering inter-landmark distances. All possible (3655) inter-landmark distances were computed and coefficients of variation were calculated for every specimen and each inter-landmark distance. The average of these coefficients was 1.3%.

Missing landmarks

Missing landmarks were estimated using the species average (separate sexes) for the missing coordinates after registration (Slice, 1999). The accuracy of missing landmark estimates was tested by simulating predictions of missing landmarks in intact specimens of *C. aethiops*.

The average number of missing landmarks per specimen with missing landmarks was < 2. Overall, the percentage of estimated landmarks (relative to the total number of 96 836 landmarks in the sample) was 0.12%.

The accuracy of the prediction was tested in *C. aethiops* including only intact specimens *sensu* Cardini & Thorington (2006). In this sample, approximately 20% of the specimens were randomly selected and the coordinates of one to six landmarks of each of these specimens deleted 'simulated' missing landmarks approximately mirrored the distribution of missing landmarks in the sample used for analysis (e.g. if L78 was missing in approximately 3% of the specimens in the original sample, L78 was deleted in approximately 3% of the intact specimens). Simulated missing landmarks were estimated using separate means of females and males. The matrix of shape distances described the phenetic relationships among all the specimens, including those with estimated missing landmarks. These were compared to the relationships described by the matrix of shape distances for the same specimens using the original landmark coordinates (intact specimens). The correlation between the matrices was ≥ 0.997. This indicates that the relative position of the specimens in the shape space is virtually identical in the two samples ('simulated' and intact). This observation is supported by a cluster analysis on the matrix of shape distances of all specimens ('simulation' and 'real' samples), which produced a phenogram where each specimen clustered with the corresponding one with estimated missing landmarks (data not shown).

The shape distance between a specimen with simulated missing landmarks and the original intact specimen measures the magnitude of shape differences due to the error in the estimate of missing landmarks. The average shape distance between any pair of specimens in the intact samples of females (or males) of *C. aethiops* measures the average magnitude of shape differences in females (or males) of this species. The ratio of the average 'intact-to-simulated' shape distance to the average shape distance was ≤ 0.11. Thus, the error of the missing landmark estimate relative to the average shape differences in *C. aethiops* was small.

In addition, a simulation was run to test the effect of missing landmark estimates on the analysis under more extreme conditions. Only half of the intact *C. aethiops* specimens were used and one to six landmarks were deleted in 50% of them. Even when calculated in this way, the correlation between shape distance matrices of samples with simulated and intact specimens was ≥ 0.992 and the ratio of shape distances ('intact-to-simulated' divided by average PRD of a species) was ≤ 0.12.

The error in the size of specimens with estimated missing landmarks was, in all simulations, less than 1 mm (maximum absolute difference between estimated and measured size). This was less than 0.4% of the centroid size of the smallest specimen in the *C. aethiops* sample.

The percentage of specimens with missing landmarks in a species (separate sexes) was generally smaller than in any of the *C. aethiops* simulations. Thus, the error introduced by estimating missing landmarks should not exceed that measured in the *C. aethiops* simulations. Examination of PC scatterplots and phenograms (separate sexes) indicated that, across all species, none of the specimens with estimated missing landmarks was an outlier.

The inclusion of specimens with few missing landmarks allows sample size to be increased. Since the number of estimated missing landmarks was very small, this had no appreciable effects on the analysis.