FULL LENGTH ARTICLE

The Anatomical Record
Advances in Integrative Anatomy and Evolutionary Biology

WILEY

Can morphotaxa be assessed with photographs? Estimating the accuracy of two-dimensional cranial geometric morphometrics for the study of threatened populations of African monkeys

Andrea Cardini[1,2] | Yvonne A. de Jong[3] | Thomas M. Butynski[3]

[1]Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Modena, Italy

[2]School of Anatomy, Physiology and Human Biology, The University of Western Australia, Crawley, Western Australia, Australia

[3]Eastern Africa Primate Diversity and Conservation Program and Lolldaiga Hills Research Programme, Nanyuki, Kenya

Correspondence
Andrea Cardini, Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Via Campi, 103, 41125 Modena, Italy.
Email: alcardini@gmail.com; andrea.cardini@unimore.it

Funding information
SYNTHESYS; Leverhulme Trust

Abstract

The classification of most mammalian orders and families is under debate and the number of species is likely greater than currently recognized. Improving taxonomic knowledge is crucial, as biodiversity is in rapid decline. Morphology is a source of taxonomic knowledge, and geometric morphometrics applied to two dimensional (2D) photographs of anatomical structures is commonly employed for quantifying differences within and among lineages. Photographs are informative, easy to obtain, and low cost. 2D analyses, however, introduce a large source of measurement error when applied to crania and other highly three dimensional (3D) structures. To explore the potential of 2D analyses for assessing taxonomic diversity, we use patas monkeys (*Erythrocebus*), a genus of large, semi-terrestrial, African guenons, as a case study. By applying a range of tests to compare ventral views of adult crania measured both in 2D and 3D, we show that, despite inaccuracies accounting for up to one-fourth of individual shape differences, results in 2D almost perfectly mirror those in 3D. This apparent paradox might be explained by the small strength of covariation in the component of shape variance related to measurement error. A rigorous standardization of photographic settings and the choice of almost coplanar landmarks are likely to further improve the correspondence of 2D to 3D shapes. 2D geometric morphometrics is, thus, appropriate for taxonomic comparisons of patas ventral crania. Although it is too early to generalize, our results corroborate similar findings from previous research in mammals, and suggest that 2D shape analyses are an effective heuristic tool for morphological investigation of small differences.

KEYWORDS
anatomical landmarks, crania, measurement error, patas monkey, Procrustes shape, variance–covariance

1 | INTRODUCTION

Integrative taxonomy is one of the most promising set of tools to accurately assess variability in relation to classification (Dayrat, 2005). Genetics might provide strong evidence of evolutionary distinctiveness, but it offers no guarantee for either taxonomic accuracy or robustness. This is especially evident in lineages that occupy a "gray area" of evolutionary divergence. As Zachos (2016, pp. 4-5) states, "When speciation is considered a continuous process through time, the exact point at which it is considered to be complete (two species) is not key to an understanding of the whole process any more … species delimitation in practice is the imposing of a binary taxonomic concept (species or no species) on a continuous process and a continuous organismic world with vague or fuzzy boundaries." Indeed, the literature abounds with examples of problematic and unstable taxonomies with not even genetic data providing conclusive, robust and stable answers. In our own work on mammals, we have encountered many cases of this type of unresolved taxonomic conundrum. Examples are the hoary marmot (*Marmota caligata*) species complex (Cardini, 2003; Kerhoulas et al., 2015), nictitans monkeys group (*Cercopithecus* [*nictitans*]) (Butynski & De Jong, 2020; Kingdon, 2013a), baboons (*Papio* spp.) (Zinner et al., 2013; Zinner et al., 2018), and red colobus monkeys (*Piliocolobus* spp.) (Oates & Ting, 2015; Ting, 2008), as well as nanger gazelles (*Nanger* spp.) (Chiozzi et al., 2014; Ibrahim et al., 2020; Senn et al., 2014) and dik-diks (*Madoqua* spp.) (De Jong & Butynski, 2017; Kingdon, 2013b). Molecular evidence, in fact, reveals only part of an often complex evolutionary picture and many other sources of information (e.g., morphology, behavior, ecology) should be integrated with DNA data for a deeper understanding of the history of a lineage.

In this context morphometrics, and especially its modern, version geometric morphometrics (GMM), offers a fairly simple and relatively low-cost approach to further explore group differences. Costs are particularly low when taxonomic analyses using GMM can take advantage of the extensive collections available in natural history museums and/or rely on a network of scientists to provide data, which can be done by sharing standardized photographs and/or measurements, or even by obtaining new specimens of rare populations with the help of field biologists who opportunistically collect skulls of dead animals (e.g., Chiozzi et al., 2014; Nowak et al., 2008).

Two dimensional (2D) GMM is particularly suitable to these types of exploratory analyses. This is because it only requires conventional photographs of crania or other anatomical structures on which to take measurements.

Compared to three dimensional (3D) data, 2D GMM is faster but inevitably less accurate whenever applied to a structure which is not flat (Álvarez & Perez, 2013; Cardini, 2014; Roth, 1993). Thus, especially in mammals, 2D GMM virtually always introduces an important source of measurement error (ME) by flattening the third dimension (reviewed by Cardini & Chiapelli, 2020). There are, however, expedients which may help mitigate this bias. It is generally best to demonstrate, in a subsample representative of the variation in the taxonomic group one intends to study, that the 2D to 3D (TTD) approximation is good. This is crucial when the differences being investigated are likely to be small, as typical of microevolutionary studies or those at the boundary between micro- (i.e., intraspecific) and macro- (i.e., supraspecific) evolution (Cardini, 2014; Cardini & Chiapelli, 2020).

1.1 | The case of the patas monkey

Well resolved and widely accepted taxonomies are important to our understanding of evolution, ecology, and behavior, and to the management and conservation of species and subspecies (Cotterill et al., 2014; Grubb et al., 2003; Zachos, 2016). This "taxonomic ideal" has, however, seldom been attained. For most taxonomic groups, a start has been made, but much work remains to be undertaken. One example of the magnitude of this problem is the current taxonomy for Africa's primates with its many on-going debates and the endless list of unanswered questions (Butynski et al., 2013; Groves, 2001; Grubb et al., 2003). A more specific example is provided by the patas monkeys (genus *Erythrocebus*), a group of large, slender, long-limbed, semi-terrestrial guenons, endemic to tropical Africa (Isbell, 2013). The geographic distribution of patas monkeys (hereafter referred to as "patas") is large, extending from Senegal across the Sahel to Tanzania (Figure 1 and De Jong & Butynski, 2020a; De Jong et al., 2020). About 19 patas taxa have been described since 1775 (Elliot, 1913; Groves, 2001; Hill, 1966; Napier, 1981).

The current taxonomy for patas, as recognized by IUCN (2020), comprises three species and three subspecies: western patas (*E. patas patas*), Aïr Massif patas (*E. patas villiersi*), eastern patas (*E. patas pyrrhonotus*), Blue Nile patas (*E. poliophaeus*), and southern patas (*E. baumstarki*). This taxonomy is based largely on the study of the colouration and pattern of the pelage. There are only meager morphological data to support this taxonomic arrangement and no molecular data. For all six taxa, the historic geographic limits are poorly known. In terms of the survival of these taxa, the *IUCN Red List of Threatened Species* (IUCN, 2020) indicates that western patas is "Near
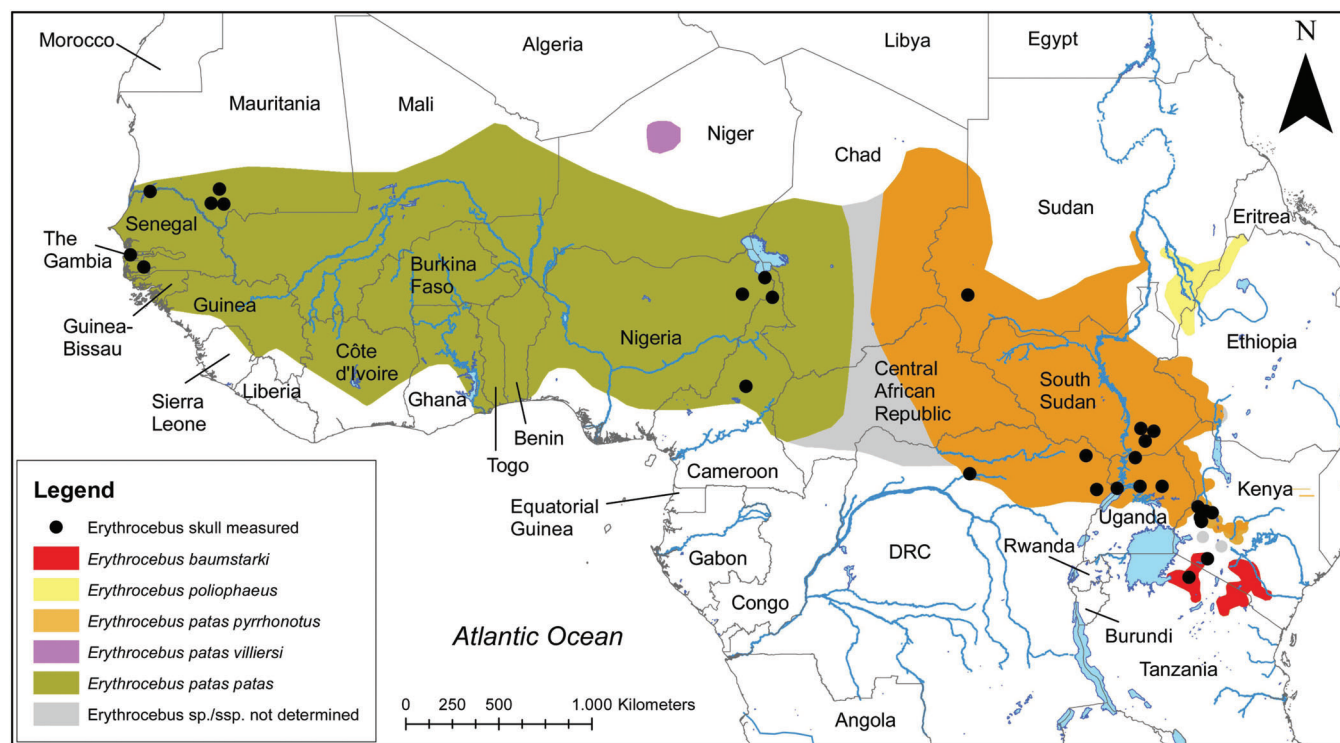
FIGURE 1 Geographic distribution of the patas monkeys (*Erythrocebus* spp.) with 29 locations depicted for those crania used in this study for which provenance is known. Provenance not accurately known for seven crania: Three from zoos, two *E. baumstarki*, and two *E. p. pyrrhonotus*. Map based on De Jong et al. (2020) and De Jong and Butynski (2020a, b; 2021)

Threatened" (Wallis, 2020), eastern patas is "Vulnerable" (De Jong & Butynski, 2020b), Blue Nile patas is "Data Deficient" (Gippoliti & Rylands, 2020), and southern patas is "Critically Endangered" (De Jong & Butynski, 2020a). The degree of threat status for Aïr Massif patas has yet to be assessed for the Red List.

Patas taxonomy is contentious and in need of support or revision, an exercise to which the three authors of this article are contributing. Since some phenotypic characters have already been examined, at least preliminarily, progress toward this end is largely dependent upon detailed morphological, ethological, biogeographical, and molecular studies. Through the research presented in this article, we hope to move one step closer toward this goal as we prepare to undertake a geometric morphometric study of the crania of patas obtained from across its geographic distribution. In the longer term, however, we hope to complement the quantitative study of morphological differences with multiple lines of evidence using an integrative approach.

1.2 | Objectives

In this paper, we examine TTD by taking advantage of an available dataset of 3D landmarks collected on

crania obtained for a previous project on the morphological variability of African monkeys (Cardini & Elton, 2017, and references therein). This dataset includes a sample of patas. The sample is small (Figure 1; Table 1) because this genus, which occurs naturally at low densities over most of its range, is poorly represented in museum collections. Nevertheless, this sample is precious, as it includes most of the complete adult specimens found in some of the largest museums of North America and Europe, and because we have 2D photographs for the same individuals for which we have 3D data. To explore whether 2D GMM provides a promising tool for studying the problematic taxonomy of patas and to obtain important information for conservation, we applied 2D and 3D cranial landmarks to the same adult individuals in order to:

1. estimate digitizing error in the 2D photographs and use this estimate as a "benchmark" to better understand the impact on size and shape of a typically much larger source of ME such as the TTD approximation;

2. assess the magnitude of TTD ME relative to the biological variability in size and shape;

3. investigate the nature of the TTD error by exploring the variance–covariance structure of shape data;

TABLE 1 Sample composition for western patas monkey (*Erythrocebus patas patas*), eastern patas monkey (*E. patas pyrrhonotus*), and southern patas monkey (*E. baumstarki*)

Sex	Eastern	Southern	Western	Zoo	Total
Females	4	1	5	2	12
Males	15	3	5	1	24
"Sex-corrected"	19	4	10	3	36

4. perform parallel tests in 2D and 3D using "biological questions", including "taxonomic questions" in relation to group differences in shape and allometry, to test whether 2D data provide the same answers as the more accurate 3D landmarks.

2 | MATERIALS AND METHODS

2.1 | Data: Samples and measurements

2.1.1 | Sample and landmarks

Collection localities are shown in Figure 1 and sample composition is detailed in Table 1. All 36 specimens are adult and all but three are wild animals. The three individuals from zoos do not show unusual cranial morphologies compared to wild specimen and are included in this methodological study to increase sample size (N) and statistical power. As mentioned, obtaining large samples of patas is challenging as they are uncommon in museum collections. This implies that many specimens can be measured only by visiting a large number of museums, which inevitably requires much time and money. Individuals used in this study belong to the collections of the American Museum of Natural History (New York), United States National Museum (Washington, DC), Harvard Museum of Comparative Zoology (Cambridge), British Museum of Natural History (London), Powell Cotton Museum (Birchington, UK), Museum für Naturkunde (Berlin), and Royal Museum for Central Africa (Tervuren, Belgium). A more detailed description of the sample and a list with catalog numbers is available at: www.wildsolutions.nl/morphometrics. Most of the 3D data used in this study are part of a larger set of data on guenons (Tribe Cercopithecini) analyzed by (Cardini et al., 2007; Cardini & Elton, 2008a, 2008b, 2017), described in (Cardini & Elton, 2017), and available for download at: http://www.italian-journal-of-mammalogy.it/Is-there-a-Wainer-s-rule-Testing-which-sex-varies-most-as-an-example-analysis-using,78976,0,2.html).

The configuration of landmarks on the ventral side of the cranium is shown in Figure 2 and described in Table 2. These also detail which 3D landmarks of the original configuration used in previous studies they correspond to. The ventral cranium was chosen as it has a fairly large number of almost coplanar landmarks, which aids in reducing the error due to the flattening of the third dimension. The ventral view of the cranium is also relatively easy to position in a standardized orientation, so that it lays approximately parallel to the lens of the camera (Panasonic Lumix DMC-TZ6 with Leica lens, held ca. 35–45 cm from the specimen, with either no zoom or a maximum 2× zoom factor and resolution of 1,984 by 1,488 pixels). We selected those anatomical landmarks that are both available in this region in the 3D dataset and easy to see in the photographs. A few of the landmarks on the alveolar margin of the premolar-molar toothrow were, however, excluded to reduce redundancy and to have a more evenly spread set of landmarks. From the total configuration of 25 landmarks, we also selected three smaller ("reduced") configurations that might help to improve the TTD approximation and compared results from these subsets to those of the full configuration. To decide which landmarks to include in the reduced configurations, we considered both the putative importance of the information they convey and estimates of ME (see below for more details). Once we found which of the reduced configuration datasets had a smaller TTD error (see point 2 of the section on statistics), we performed all further analyses (3–4) on this reduced set of landmarks, as well as on the full configuration.

2D landmarks were digitized in TPSDig 2.31 (Rohlf, 2015), only on the left side of the cranium. Although it is generally better to landmark and measure both sides of a symmetric structure (Cardini, 2017), we opted for a one-side only approach for consistency with the 3D dataset. For the 3D data (collected using a three-dimensional digitizer (MicroScribe 3DX; Immersion Corporation, San Jose, CA), this also reduced costs and maximized sample size. The left-side only 2D-measuring of landmarks was repeated after 1 week to have replicates to estimate digitization error. Both sets of measurements were taken by AC. Paired landmarks were then mirrored and the very small asymmetries on the midplane removed following (Cardini, 2017). Because the photographs of the ventral crania were originally taken as snapshots with no specific research purpose, the scale factor in those images is approximate and may show small inconsistencies among individuals. Thus, as in
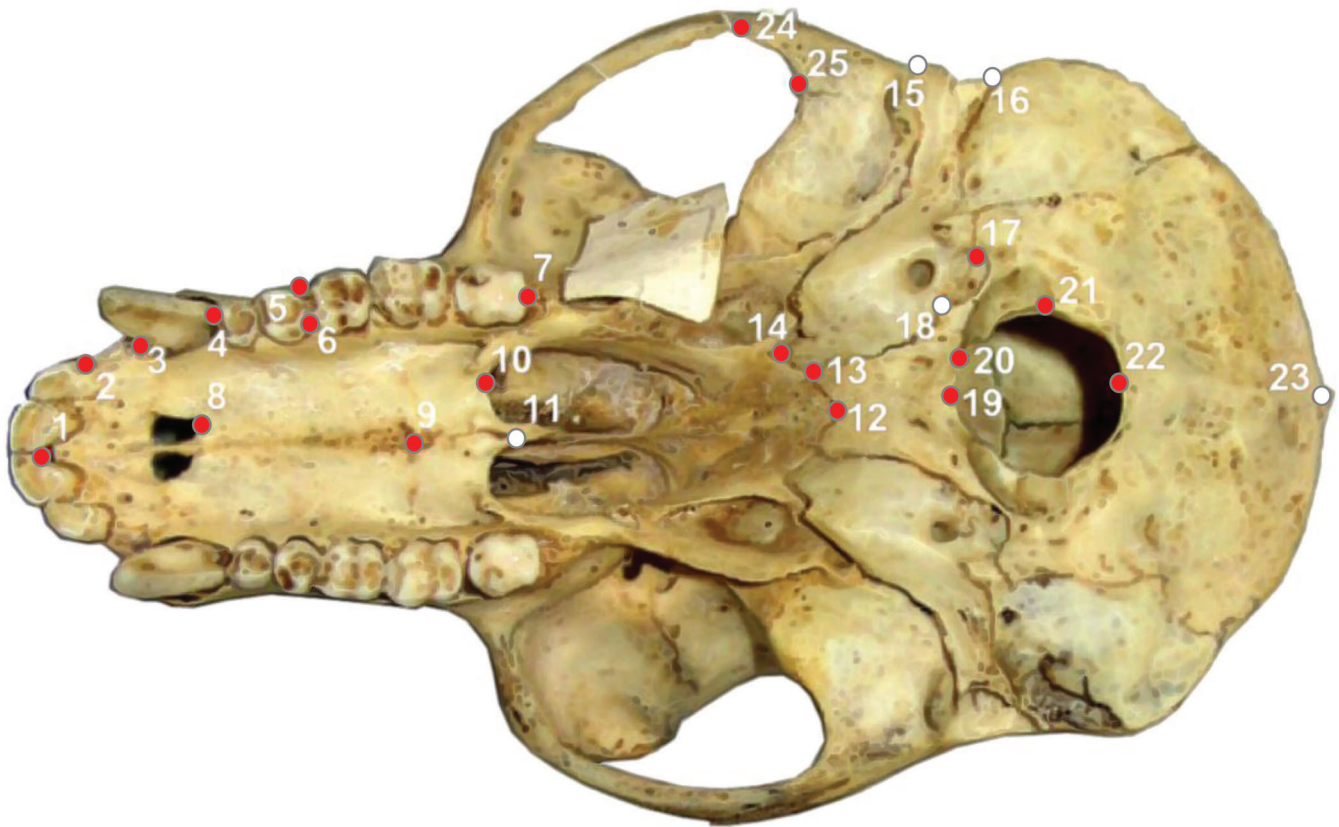
FIGURE 2 Landmark configuration (FULL). Red landmarks are those of the reduced configuration with the smallest TTD ME (i.e., RED-4)

Cardini and Chiapelli (2020), the 2D data were rescaled using a measure of condylobasal length obtained from the 3D coordinates between landmarks 1 and 19. This is akin to using a simple but accurate caliper measurement taken directly on the crania instead of a distance measured using a scale factor in the photograph, as customary in 2D GMM analyses.

2.1.2 | Geometric morphometrics and software

Landmark-based GMM extracts the features of interest from a study structure by computing the centroid size of a set of anatomical landmarks and, having set centroid size to unit in all individuals, by standardizing positional differences in the sample using a Procrustes superimposition (Rohlf & Slice, 1990). The superimposition produces a new set of variables, the Procrustes shape coordinates. This allows the measurement of multivariate differences in terms of Procrustes distances, which are generally equivalent to Euclidean distances in a multidimensional Euclidean space tangent to the curved Procrustes space (Rohlf, 2000). It is, however, important to bear in mind,

as stressed in Cardini (2020a), that Procrustes is a least square approach, which has statistically desirable properties but is not based on a biological model. The "biological arbitrariness" of this choice, thus, prevents univariate analyses of the shape variables, as well as analyses and interpretations of variance one landmark at a time. In contrast, this method performs well, and results can be accurate, when shape coordinates are analyzed all together using multivariate methods (Rohlf, 1998) and the findings are carefully interpreted using diagrams that integrate patterns of covariation over the entire set of landmarks (Klingenberg, 2013).

Another consequence of the Procrustes superimposition is that shape spaces are specific to each landmark configuration. This is also true when the landmarks are the same but one set is 2D and the other is 3D. The dimensionality of the data is also necessarily different in this case, because 2D data have only two coordinates (*X* and *Y*) for each landmark, whereas 3D data have a third (*Z*) coordinate. For these reasons, 2D and 3D results are usually compared with correlational analyses in separate data spaces (as we do specifically to answer our fourth study question). The two types of data can, however, be "brought" into the same space using an

TABLE 2   Landmark definitions: Gray background shows potentially problematic landmarks (underscored for those with largest digitizing error in 2D; bold for those with largest difference between 2D and 3D shapes in the common shape space)

Midplane	Paired	Definition	Cardini et al. (2007)
1		Prosthion: antero-inferior point on projection of premaxilla between central incisors	1
	2	Posterior-most point of lateral incisor alveolus	3
	3	Anterior-most point of canine alveolus	4
	4	Mesial P3: most mesial point on P3 alveolus, projected onto alveolar margin	5
	5–6	Contact points between adjacent premolar/molar projected onto alveolar margin	7–13
	7	Posterior midpoint onto alveolar margin of M3	10
	8	Posterior-most point of incisive foramen	15
9		Meeting point of maxilla and palatine along midline	16
	10	Point of maximum curvature on the posterior edge of the palatine	18
11		Tip of posterior nasal spine	19
12		Meeting point between the basisphenoid and basioccipital along midline	20
	13	Meeting point between the basisphenoid, basioccipital and petrous part of temporal bone	21
	14	Most medial point on the petrous part of temporal bone	22
	15–16	Anterior and posterior tip of the external auditory meatus	25–26
	17–18	Distal and medial extremities of jugular foramen	28, 30
19		Basion: anterior-most point of foramen magnum	31
	20–21	Anterior and posterior extremities of occipital condyle along margin of foramen magnum	32, 35
22		Opisthion: posterior-most point of foramen magnum	36
23		Inion: most posterior point of the cranium	37
	24	Zygo-temp inferior: infero-lateral point of zygomaticotemporal suture on lateral face of zygomatic arch	55
	25	Posterior-most point on curvature of anterior margin of zygomatic process of temporal bone	56

Note: The last column shows the corresponding landmarks in previous studies.

expedient (Cardini, 2014; Cardini & Chiapelli, 2020). This simply involves adding a zero Z coordinate to the X and Y coordinates of the 2D data, superimposing these data with the 3D data (Figure 3a), and mean-centering the two sets of shapes to remove the bias due to the missing information in the third dimension (Figure 3b). This allows for exploratory analyses in the same data space, such as ordinations and phenograms (Viscosi & Cardini, 2011), as well as analysis of variance (ANOVA; Klingenberg et al., 2002), which is conventionally used to assess whether individual variation in a sample is larger than ME.

For this study, data were Procrustes superimposed and mostly visualized in MorphoJ 1.07a (Klingenberg, 2011), although we did most of the statistical analyses in R (R Core Team, 2020) using scripts written by AC. Specifically, we used the following main R packages: *vegan* (Oksanen et al., 2011) for permutational analyses of
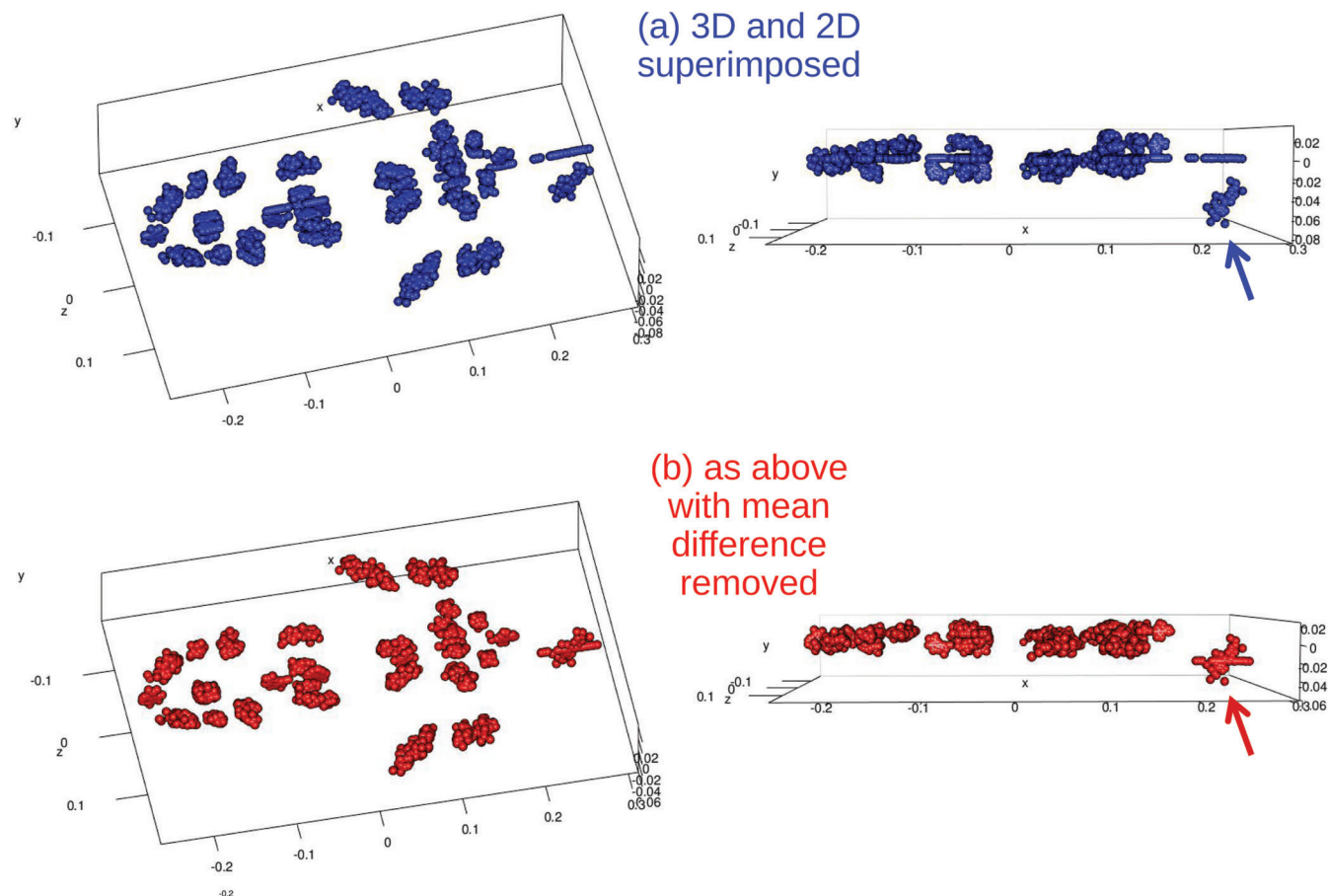
(a) 3D and 2D superimposed

(b) as above with mean difference removed

FIGURE 3 Procrustes superimposed 3D and 2D shapes (FULL-1) before (a) and after (b) removing their mean difference (side view in the right half of the figure). The arrows indicate inion (landmark 23), which is off the main plane of the other landmarks even after mean-centering

variance and covariance, and for matrix correlations; *car* (Fox & Weisberg, 2011) and *adegraphics* (Siberchicot et al., 2017) for ordinations and scatterplots; *MASS* (Venables & Ripley, 2002) for simulating random normally distributed multivariate data; and *shapes* (Dryden, 2019) and *Morpho* (Schlager, 2017) for analyses which required that the superimposition be redone for visualizing 3D superimposed shapes in a common space, and for testing classification accuracy according to taxonomic groups.

2.1.3 | Dataset abbreviations and criteria for selecting reduced configurations

A summary of all the main abbreviations used in this study is presented in Appendix A. Here, we describe, in more detail, the abbreviations for the different sets of data:

"FULL" is the total configuration of 25 left side landmarks (Figure 2), which became 43 after mirror reflection of the paired landmarks. "FULL-0" refers to the data

using the full configuration in the two 2D replicates (i.e., the same photographs digitized twice). "FULL-1" is the same configuration both with 2D (using for each individual the mean of the two digitizations) and 3D data.

"RED" refers to the reduced configurations where landmarks potentially affected by a relatively larger ME were removed. To decide these potentially "problematic" landmarks we used three criteria:

○ "RED-2": in this configuration, we excluded landmarks with the largest 2D median digitization error in FULL-0. These were selected by summing the variance of X and Y coordinates of, for instance, the two replicates of landmark 1. For this, we used the raw coordinates (before doing any superimposition) because they are separate digitizations of the same photograph and, therefore, any difference in X and Y is purely due to digitizing error. As an estimate of the average digitizing error of this landmark, we took the median of the variances of landmark 1 across all 36 individuals. We did the same for all landmarks. Finally, we excluded

the anatomical landmarks whose median 2D digitizing error is larger than the 90th percentile of the medians of all 25 landmarks. Specifically, we removed landmarks 11, 15, 16 and 18 (and the corresponding mirror-reflected paired landmarks). In fact, landmark 16 had a variance (12.5 mm^2), which is slightly lower than the 90th percentile (12.8 mm^2), but we decided not to include it because we were already aware that 15 and 16 (anterior tip and posterior tip of the external auditory meatus) are difficult to locate precisely both in 2D and 3D.

○ "RED-3": in this configuration, landmarks with the largest median 2D-3D shape variance in FULL-1 were removed. They were selected using the same procedure as in RED-2 but this time the first replicate was 3D data and the second replicate was 2D (means of the two digitizations on the photographs) brought into the same shape space, superimposed (Figure 3a) and mean-centered (Figure 3b) (Cardini, 2014). RED-3 tentatively explores whether a specific landmark might have a very poor TTD approximation. The hints provided by this analysis, however, should be taken with great caution and are potentially misleading because, as mentioned, after a Procrustes superimposition, landmarks cannot be analyzed or interpreted one at a time (Rohlf, 1998; Viscosi & Cardini, 2011, and references therein). Using this approach we excluded landmarks 7, 11 and 23. As this information is preliminary to the main analyses, we need to mention that we did not expect the posterior extremity of the premolar-molar toothrow (landmark 7) to have a particularly large error. This might be a case where we were misled by the superimposition spreading the shape variance across the whole set of landmarks (thus, potentially inflating or deflating variance locally). In contrast, for the other two landmarks (11 and 23), we believe the results are reliable, as we had anticipated that they might have a very large ME. The posterior tip of the nasal spine (11) is difficult to locate precisely, which is why it also has a large 2D digitization error and was excluded from RED-2. The inion (23), however, is relatively easy to digitize with a good replicability but, as suggested by Figure 3a (showing the superimposed 2D and 3D data before mean-centering), it tends to lie well off the plane of most other landmarks and is, therefore, strongly affected by TTD distortions.

○ "RED-4" (marked in red in Figure 2): in this last configuration, we selected landmarks by complementing clues obtained in the selection of RED-2 and RED-3 with our knowledge of anatomy to decide which landmarks might be particularly problematic or informative. We, therefore, excluded all those of RED-2 and RED-3 which we most expected to have large errors, but made an exception for landmark 7 (the posterior midpoint onto the alveolar margin of the third molar).

We decided to keep this landmark since we are skeptical about its apparently large TTD error (as explained above), and also because measuring the length of the masticatory toothrow potentially provides information on a taxonomically and functionally important trait.

2.2 | Statistics

Before detailing the statistical analyses, which we subdivide into four main subsections corresponding to the four sets of research questions, we here clarify more precisely what components of ME we have assessed, as well as why we believe that flattening the third dimension is the main reason for ME in 2D data compared to 3D (i.e., in FULL-1 and the three RED configurations). For additional information on different sources of ME in GMM, please refer to the reviews of Arnqvist and Martensson (1998) and Fruciano (2016), as well as a recent case study by Fox et al. (2020).

Because all data were collected by a single operator, there is no interoperator bias, which can be a large contributor to ME in GMM shape data (Daboul et al., 2018; Fox et al., 2020; Fruciano et al., 2017). In the newly collected 2D data, however, besides digitizing error (i.e., the precision or repeatability in locating the landmarks in the same photograph of a specimen), which we assessed, there could be a further source of error in relation to how well (or poorly) the orientation of a specimen in a photograph can be standardized. In fact, orientation errors, together with digitizing error, may affect both 2D and 3D data, with the additional difficulty, in 2D, of landmarking a photograph instead of doing it directly on the 3D structure. Both of these components of ME are, however, part of the differences between 2D and 3D and, therefore, indirectly included in the assessment of the TTD approximation. Nonetheless, it is likely that 2D flattening is a major source of ME in our TTD analysis. This is why we largely interpret the estimates of ME in this main part of the study as principally due to the distortion of the third dimension in the photographs. This interpretation is consistent with the finding (see Section 3) that individual shape variance is 14 times larger than 2D digitizing error (and on average more than 10 times larger using 3D skulls of guenons; Cardini & Elton, 2008a), whereas in the TTD comparison it was only four to five times bigger. Thus, if the digitizing error in the TTD comparison is only slightly larger than in 2D, with the orientation error typically about as large as digitization error (smaller sometimes: Evin et al., 2020; Fox et al., 2020; Jojić et al., 2011; Souto et al., 2019; Jojic, pers. comm.; Cardini, unpublished; although larger in a few cases: Fruciano, 2016; Klenovšek & Jojić, 2016; Murta-Fonseca et al., 2019), the flattening of the third dimension becomes the most likely main source of TTD error.

2.2.1-2.2.2 | Analyses comparing the magnitude of ME to biological differences in size and shape using ANOVAs (1) and correlational/graphical approaches (2)

We assessed whether the magnitude of variation among averaged replicates of the individuals in the sample (which for brevity we refer to as "biological" variation) is larger than differences between replicates using hierarchical ANOVAs (analysis of variance) with sex as the main factor and individuals as a random factor (Fruciano, 2016; Klingenberg et al., 2002; Viscosi & Cardini, 2011). This design controls for sex differences before comparing individual variation to the residual variance, which represents differences between replicates (i.e., our estimate of ME). Thus, by statistically controlling for sexual dimorphism, we avoid underestimating the importance of ME relative to individual variation. Taxonomic comparisons of appreciably dimorphic taxa using separate analyses for females and males are generally more desirable (Cardini, 2020a) than "sex-corrections" (such as the one we used here). However, we adopted the strategy of statistically controlling for sex differences in order to increase N in a study where N is low and which is mostly methodological. Thus, as our main aim is the assessment of ME, we preferred to tolerate the cost of a small reduction in biological accuracy in order to gain a higher statistical power.

In relation to sex-correction, we concisely report here an issue that has no consequences on the robustness of our results. In the specific case of an ANOVA on 2D and 3D shapes, brought into the same shape space using Cardini's (2014) approach, the effect of sex is not completely removed if some of the analyses are later performed separately on the two types of data. For instance, by retesting sex within sex-corrected 2D (or 3D) shape alone, one finds that, on average, about 3% of variance is still explained by sex. This small effect (about 10 times smaller than the total observed sex differences in shape—see Section 3) is a consequence of the interaction between sex and type of data (2D vs. 3D). This interaction term was not included in the ANOVA model because it is small and not significant. If included, however, the sex-correction completely removes sex differences, as if an ANOVA had been done separately within 2D (or 3D). The effect of the interaction in our dataset is, therefore, real but negligible. Indeed, sex-corrected shape distances, without or with the interaction in the model, have an almost perfect matrix correlation ($r = 0.98$–1.00, depending on the configuration). This indicates that the shape similarity relationships are almost the same despite the small amount of sexual dimorphism left in the residuals of the ANOVA with no interaction term. In this study, therefore, the interaction was ignored in all

analyses except the parallel tests of taxonomic differences (i.e., the fourth set of tests, described below), which are run separately in 2D and 3D. Yet, even in this last set of analyses, as in all others, the two slightly different ways of sex-correcting shapes produced virtually identical results. Even repeating the parallel tests using only the larger male sample (thus, using no sex-correction) did not appreciably change our findings (as we concisely summarize here). Using male-only data in FULL-1 and RED-4, the main differences were the typically larger percentages of variance explained in the tests of group differences (7–12%, and consistently slightly larger in 2D, although significant only in FULL-1 2D) and the marginally better prediction of taxonomic affiliation using shape in FULL-1 (75–85% cross-validated accuracy in, respectively, 3D and 2D). This first observation is expected, because R^2 tends to be over-estimated in small samples (Cramer, 1987; Nakagawa & Cuthill, 2007), while the second observation may be a consequence of a smaller within-taxon heterogeneity (and thus better separation) in same-sex data compared to sex-corrected data. However, as mentioned, none of these small differences changed the conclusions of the main analyses and are briefly discussed here so as to not further distract the reader from the main findings.

We performed ANOVA tests both on size and shape with the same design using the *vegan adonis()* function and permutations of Euclidean distances (Oksanen et al., 2011). In this and all other tests of significance (see next sections) we used 10,000 permutations for the specific test statistics. For digitizing error in 2D data (FULL-0), the first and second replicates correspond to the first and second digitization of landmarks. In contrast, for the analysis of the TTD approximation (FULL-1 and all the RED configurations), the first replicate used the 3D data and the second replicate the 2D data (averaged between the two digitizations of FULL-0). When analyses required a common shape space, as in the case of the ANOVA, phenograms, and some of the ordinations (see below), the 2D and 3D sets of data were the mean-centered shapes superimposed together, as explained in the section on GMM (Cardini, 2014).

Using "sex-corrected" data (i.e., after removing the mean differences due to sex), we computed the correlation between estimates of centroid size and shape distances in the first and second replicate. For shape, we did this by computing the matrix correlation of Procrustes shape distances in 2D and 3D. Also, for shape, we calculated the proportion of specimens with the two replicates of an individual clustering together as "sisters" in phenograms (unweighted pair group method with arithmetic mean – UPGMA – using the matrix of Procrustes sex-corrected shape distances in the common shape space). If differences between

replicates are small relative to those among specimens in a sample, the expectation is that they will cluster in pairs as nearest neighbors.

2.2.3-2.2.4 | Exploring the structure of ME (3) and the congruence of results (4) in 2D with those in 3D

In the sections dedicated to the third (2.2.3) and fourth (2.2.4) set of our research questions, we restricted the analyses to shape in the two configurations which are potentially more interesting, either because they include the most complete information (FULL) or are the most promising in terms of smaller TTD error (RED-4, see Section 3). We did not analyze size because, as in previous studies (Cardini, 2014; Cardini & Chiapelli, 2020), analyses (1–2) demonstrated that centroid size in 2D corresponds accurately to its estimates in 3D, except for a very small bias, which likely leads to a slight underestimate of 2D size. Also, unless we specify otherwise, we analyzed only sex-corrected shape in order to control for the otherwise dominant effect of sex differences.

As before, we employed the FULL configuration in the two sets of 2D digitizations (i.e., the FULL-0 dataset) as a sort of "benchmark" to compare the impact of a strong source of ME, such as the TTD approximation, to a much smaller one, such as digitizing error. Thus, section three explores covariance in relation to ME (2D digitizing error in FULL-0 or TTD error in FULL-1 and RED-4). Section 4, in contrast, explores the congruence of findings from analyses of group differences by comparing them either between the first and second replicate of the 2D landmarks (FULL-0) or between the 2D averaged (between replicates) shapes and the 3D data (FULL-1, RED-4).

2.2.3 | Structure of ME

Variance and covariance in relation to ME: Differences in variance covariance

First, using matrix correlations, we explored how well covariance in 2D corresponds, in terms of overall proportionality, to estimates in 3D. We then tested if the magnitude of shape variance is the same in 2D and 3D using paired permutation tests for estimates of total multivariate variance. The tests for differences in magnitude of shape variance are the multivariate equivalent of a (paired) Levene's test for univariate data (Willmore et al., 2006). As test statistics, we employed the absolute difference of each of three estimators of overall multivariate variance: VAR1, the sum of variances of the shape coordinates (whose test is approximately equivalent to using an F ratio—thus,

showed in Section 3 together with the corresponding R^2, but not tested); VAR2, the mean of pairwise Procrustes shape distances among all individuals in a sample; VAR3, the 90th percentile of the same set of pairwise Procrustes distances used in VAR2. VAR1 is intuitive, as it is a straightforward extension of univariate variance. VAR2 and VAR3 can be interpreted, respectively, as the average difference in shape among all individuals in a sample and the equivalent, based on multivariate distances, of a trimmed range (from minimum to maximum) for univariate data. These types of variance metrics are commonly employed in disparity analyses (Foote, 1997).

Although we did these analyses on sex-corrected shapes, as anticipated, data without sex-correction produced very similar results (not shown).

Variance and covariance in relation to ME: Exploring the strength of covariance in the ME component of shape variation

Cardini and Chiapelli (2020) found that, despite a large TTD error in ventral cranial shapes of equids, tests of "biological" questions (such as the relevance and patterns of allometry and species or sex differences) produced results in almost perfect agreement when performed in parallel on 2D and 3D data. They speculated that, despite inaccuracies in the precise relative positions of the specimens in the 2D shape space compared to the 3D shape space, and the relatively large ME (accounting for ca. 9–19% of variance within, respectively, genus *Equus* and plains zebra [*E. quagga*]), the biological covariance of the data was much stronger than that of ME. Thus, they argued that the genuine "signal" (behind the test results and patterns) could overcome the "noise" introduced by the TTD approximation.

To explore Cardini and Chiapelli's (2020) hypothesis on why 2D may produce accurate results despite an apparently fairly poor TTD approximation, we first checked the magnitude of the correlation between covariance matrices of the first and second replicate. We then compared the mean covariance of the individuals with that of ME (be it the 2D digitizing error, again used as a "benchmark," or the TTD inaccuracy). Thus, we separated the individual from the error component of shape using the same ANOVA design as in the tests of ME (1–2). On these variables, we computed the covariances of the shape data for the individuals and the ME, and summarized them using histograms and the median of the absolute covariances. The use of the absolute value is justified, as we were not interested in the sign of covariances but only in their average magnitude.

Since we found that covariance is on average much smaller in ME (see Section 3), we wondered if the observed small ME covariances might be simply random noise. This would mean that the error due to the TTD approximation, or the one related to the imprecision of

2D landmarks, is not structured. We used the *mvrnorm()* function in R (Venables & Ripley, 2002) to simulate random normally distributed data with variance and no covariance. In the simulation, variances were those observed in the ME component of shape, but covariances were set to zero. Any covariance found in the simulated data is, therefore, the product of sampling error. However, because the Procrustes superimposition introduces a degree of covariation when nonshape parameters are standardized (Cardini, 2019; Rohlf, 1998; Rohlf & Slice, 1990), we added the simulated random numbers to the mean shape of the dataset and performed a Procrustes superimposition. With these "random" shape data we computed the covariance matrix to be compared with the observed ME covariances. This extra step does not seem to make an appreciable difference as before and after the superimposition the simulated covariance matrices had a correlation of 0.95 or larger. We preferred, however, to include the superimposition of random data to increase comparability with observed shape data. This simulation was run 100 times for each dataset (i.e., FULL-0, FULL-1, and RED-4) and in each simulation we computed the median absolute covariance. The resulting 100 medians were summarized using a histogram, in which we also plotted the medians of the absolute covariances both of the observed ME and individual shape data.

2.2.4 | Parallel tests of taxonomic differences

In this final part of the study we performed a series of tests related to taxonomic differences between the two largest taxonomic samples, the western patas and eastern patas (thus, excluding the four southern patas crania and three zoo specimens crania). Two analyses were undertaken in parallel, one to compare the first and second digitizations of 2D data (assessing the impact of 2D digitizing error on results), and one to compare the 2D and 3D shape variables (the most interesting comparison to investigate if conclusions from 2D shapes are accurate). As anticipated, we did not analyze centroid size because its ME is, as determined in previous studies (Cardini, 2014; Cardini & Chiapelli, 2020), typically negligible. We did, however, briefly explore centroid size graphically (using box-plot and violin plots) to determine if 2D data really suggest the same pattern of size variation as 3D data in relation to taxonomic groups. On shape, in contrast, we performed three sets of statistical tests:

Differences in mean shapes between taxa
We tested the significance of mean differences between western patas and eastern patas in a hierarchical permutational multivariate ANOVA controlling for sex

(*adonis()* function using Euclidean distances in *vegan*; Oksanen et al., 2011). Thus, sex is the first main factor followed by taxon. We also calculated the mean classification accuracy (hit rate) of the sex-corrected shape data using a cross-validated between group principal component analysis (XbgPCA; Cardini & Polly, 2020) in Morpho (*groupPCA()* function; Schlager, 2017) and assessed if it was better than expected by chance using a random chance baseline based on 100 randomizations (Kovarovic et al., 2011; Solow, 1990; White & Ruttenberg, 2007). Finally, we illustrated the patterns of sex-corrected shape variation in the two taxa using ordinations (principal component analysis, PCA, and also XbgPCA) and visualized the mean differences in shape with wireframe diagrams both for 2D and 3D data (Klingenberg, 2013).

Differences between taxa in magnitude of sex-corrected shape variance
These were tested using permutations and the same test statistics (VAR1, VAR2, and VAR3) as explained in the previous subsection but, because the groups (i.e., western patas and eastern patas) are now independent, the permutations simply affiliated randomly the individuals to the groups to simulate the null hypothesis of no differences.

Static allometric trajectories of sex-corrected shapes in western patas and eastern patas
We tested the significance and divergence of static allometries (Klingenberg, 1998) using a permutational multivariate analysis of covariance (ANCOVA, in *vegan* using the *adonis()* function and Euclidean distances; Oksanen et al., 2011). In this analysis, the "taxon by centroid size" interaction tests the significance of the differences in the allometric trajectories, for which we also computed the angles in the multivariate space.

3 | RESULTS

As with the methods, we present the results by subdividing them into four sections which correspond to the four sets of research questions.

3.1-3.2 | Analyses comparing the magnitude of ME to biological differences in size and shape

ANOVAs of size (Table 3) are dominated by sex differences that in all datasets account for about 75% of total variance. Individual variation, sex-corrected by controlling for differences between females and males, accounts for almost all the remaining variance (25%), with ME

TABLE 3 Ventral cranial size: analysis of variances testing if individual variation is significantly larger than measurement error (ME); in the last column, the correlation between the two replicates, after controlling for sex, is also shown

Configuration	Factor	df	SS	MS	F	p	R^2	R^2 ratios	r
FULL-0	Sex	1	57,869.5	57,869.5	101.53	.0001	74.9%	3	
	Individual	34	19,379.5	570.0	614.57	.0001	25.1%	580	1.00
	ME[a]	36	33.4	0.9			0.04%		
	Total		77,282.3				100.0%		
FULL-1	Sex	1	59,202.8	59,202.8	101.56	.0001	74.7%	3	
	Individual	34	19,820.0	582.9	97.01	.0001	25.0%	92	0.98
	ME	36	216.3	6.0			0.3%		
	Total		79,239.1				100.0%		
RED-2	Sex	1	49,132.4	49,132.4	99.05	.0001	74.3%	3	
	Individual	34	16,865.3	496.0	111.40	.0001	25.5%	105	0.99
	ME	36	160.3	4.5			0.2%		
	Total		66,158.0				100.0%		
RED-3	Sex	1	56,164.7	56,164.7	104.69	.0001	75.3%	3	
	Individual	34	18,241.1	536.5	137.37	.0001	24.5%	130	0.98
	ME	36	140.6	3.9			0.2%		
	Total		74,546.4				100.0%		
RED-4	Sex	1	46,900.0	46,900.0	101.76	.0001	74.9%	3	
	Individual	34	15,671.0	460.9	219.40	.0001	25.0%	207	0.99
	ME	36	75.6	2.1			0.1%		
	Total		62,646.6				100.0%		

Note: In this and following tables, significant ($p < .05$). p values are emphasized using italics or bold italics if highly significant ($p < .01$); R^2 ratios compare the variance explained by one factor (numerator) to the one explained by the next factor (denominator; i.e., how large sex-related variance is compared to individual differences or how large the individual differences are compared to ME); a gray background is used for results of FULL-0, which only concern 2D digitizing error.

[a]This is just 2D digitizing error whereas in all other cases ME is mainly TTD error.

explaining only 0.04% (2D digitizing error in FULL-0) to 0.3% (TTD error in FULL-1) of centroid size differences. This means that individual "biological" variation is, respectively, almost 600 and 100 times larger than ME. RED-4 shows the best 2D approximation of 3D centroid size, with an R^2 of 0.1% (i.e., differences between 2D and 3D estimates of size are more than 200 times smaller than those among sex-corrected individuals). The correlation between 2D and 3D centroid size ranges from 0.98 (FULL-1 and RED-3) to 1.00 (FULL-0).

Sexual dimorphism is important also for shape (Table 4) but accounts for a smaller percentage of total variance (ca. 25%). Sex-corrected individual variation, in contrast, explains most of the variance (58–66%), whereas ME is almost two orders of magnitude larger than for size ($R^2 = 5$–15%). 2D digitizing had the smallest error ($R^2 = 5\%$) with individual variation 14 times larger than ME. TTD error is larger than 2D digitizing error, but about the same in all configurations ($R^2 = 15\%$) except RED-4 ($R^2 = 14\%$). For this reason, as well as for its

slightly larger differences among individuals compared to other configurations except RED-3, individual variation in RED-4 is five times larger than TTD ME but only four times bigger in all other configurations.

Correlations of pairwise Procrustes shape distances mirror the "ranking" suggested by the ANOVAs R^2s in terms of relative importance of ME with FULL-0 having the highest ($r = 0.93$) and RED-4 the second highest ($r = 0.80$) correlations. The other three TTD datasets (FULL-1, RED-2, and RED-3) shows lower correlations ($r = 0.71$–0-79) and, thus, confirm the marginally smaller TTD ME of RED-4. In the phenograms (not shown, except for FULL-1, used as an example in Figure 4), however, RED-4 turns out not to be the configuration with the highest percentage of correctly paired replicates. Both RED-3 and RED-2 perform slightly better than RED-4 (respectively, 50–47% vs. 44%), whereas FULL-1 performs slightly worse (42%). 2D digitizing error (FULL-0), in contrast, is again much smaller than TTD error in terms of individuals with sister replicates in the phenogram, as

TABLE 4 Ventral cranial shape: Multivariate analysis of variances testing if individual variation is significantly larger than measurement error (ME); in the second to last column, the matrix correlation between the pairwise sex-corrected shape distances (first vs. second replicate) is shown; the last column reports the percentage of cases with replicates clustering as pairs "within" individuals ("sister replicates")

Configuration	Factor	df	SS	MS	F	p	R^2	R^2 ratios	Matrix r	Sister repl.
FULL-0	Sex	1	0.05605	0.001367	14.97	*.0001*	29.1%	0.4		
	Individual	34	0.12732	0.000091	30.23	*.0001*	66.2%	14	0.93	
	ME[a]	36	0.00892	0.000003			4.6%			97%
	Total		0.19229				100.0%			
FULL-1	Sex	1	0.05449	0.000447	16.00	*.0001*	27.2%	0.5		
	Individual	34	0.11579	0.000028	4.10	*.0001*	57.8%	4	0.77	
	ME	36	0.02989	0.000007			14.9%			42%
	Total		0.20017				100.0%			
RED-2	Sex	1	0.05581	0.000553	16.28	*.0001*	27.7%	0.5		
	Individual	34	0.11655	0.000034	4.24	*.0001*	57.8%	4	0.79	
	ME	36	0.02912	0.000008			14.5%			47%
	Total		0.20148				100.0%			
RED-3	Sex	1	0.03522	0.000320	11.50	*.0001*	21.6%	0.3		
	Individual	34	0.10413	0.000028	4.61	*.0001*	63.8%	4	0.71	
	ME	36	0.02392	0.000006			14.7%			50%
	Total		0.16327				100.0%			
RED-4	Sex	1	0.04786	0.000488	13.60	*.0001*	24.7%	0.4		
	Individual	34	0.11971	0.000036	4.80	*.0001*	61.7%	5	0.80	
	ME	36	0.02638	0.000007			13.6%			44%
	Total		0.19395				100.0%			

[a]This is just 2D digitizing error whereas in all other cases ME is mainly TTD error.

this happens almost 100% of the time compared to the ca. 50% that we report above for TTD data.

Thus, among the configurations used to assess TTD, differences in the magnitude of ME are small but, both for size and shape (with the exception of the phenograms), RED-4 seems slightly less impacted by TTD errors. The next two series of analyses are, therefore, performed using RED-4, as well as the complete set of landmarks (FULL-1, plus FULL-0 for 2D digitization error).

3.3 | Variance and covariance in relation to ME

3.3.1 | Differences in variance covariance

The correlation between variance covariance matrices (Table 5) is close to one (0.95) for FULL-0 (first vs. second replicate of the 2D digitizations), but smaller (ca. 0.8) when 2D is compared with 3D (FULL-1 and RED-4). This is consistent with correlations of Procrustes shape distances and the ANOVA results (Table 4) which show a small 2D

digitization error (FULL-0) but a larger TTD error (FULL-1 and RED configurations). RED-4, again, does marginally better than FULL-1 ($r = 0.81$ and 0.75, respectively).

The difference in the magnitude of variance between the first and second digitization of the 2D photographs is significant only for VAR2 ($.05 > p > .01$) and marginally significant ($.1 > p > .05$) for VAR1, with absolute deviations from the means of the two replicates accounting for less than 1% of variation (Table 5). Thus, it seems that for FULL-0, variance is approximately the same in the two replicates and differences are overall minor and negligible. In contrast, all test statistics are significant (with most highly significant—$p < .01$) when 2D is compared with 3D using FULL-1 and RED-4. R^2 is approximately 5%, and 3D shapes consistently show larger variance than 2D.

3.3.2 | Exploring the strength of covariance in the ME component of shape variation

The correlation between covariance matrices (excluding the diagonal with variances) of individuals and ME is

WILEY

very modest (ca. 0.2–0.3; Table 6). This suggests large differences in covariance structure between these two components of shape. Absolute covariances of individuals are also typically larger than those of ME (Figure 5). This is

particularly evident when individuals are compared to 2D digitizing error (FULL-0; Figure 5a). Most of the covariances of this type of ME (in blue) are close to zero, whereas a large proportion (82%) of covariances among individuals (in red) are larger than that. The same is true for TTD errors in FULL-1 and RED-4 (78% of individual covariances larger than those of ME), but TTD error seems to show more structure and a larger overlap with individual absolute covariances.
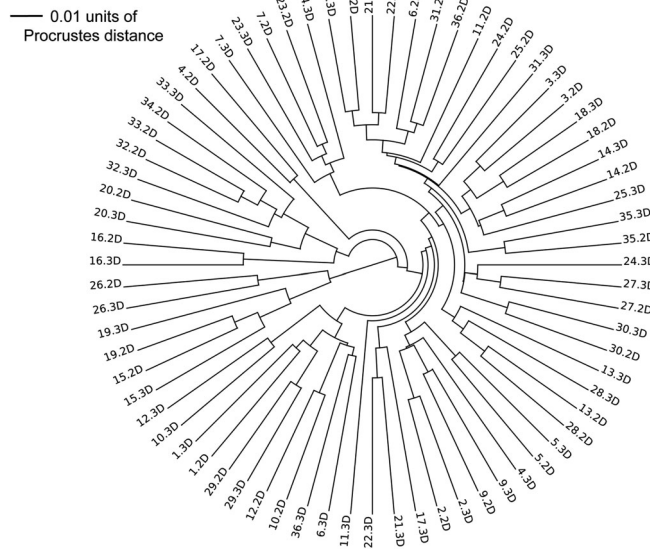
Indeed, that covariance is largely random for 2D digitizing error but more structured for TTD is confirmed in the simulation of random data with variance of the same magnitude as ME but with no covariance. As an example, Figure 6 illustrates the pattern of covariance of individuals, ME error, and simulated random noise by using a PCA performed on each of the three types of data. The example is specific to FULL-0 (with ME due to 2D digitizing error) and RED-4 (with ME being mainly that of the TTD approximation), but the reasoning and results (not shown) are similar in other cases. In general, multivariate data with a strong covariance should produce a few dominant PCs, as a small number of dimensions captures most of the variance in highly correlated data. This is what happens for the individuals in both datasets (Figures 6a1,b1). In contrast, random noise should produce PCs which explain about the same amount of variance, although a certain degree of "structure" might be found even for random noise when sample size is small compared to the number of variables (Bookstein, 2019;

FIGURE 4 Phenogram using, as an example, the sex-corrected FULL-1 Procrustes shape distances. Individuals are numbered progressively from 1 to 36. This number is followed by a label indicating the type of shape data (2D or 3D). The tree shows 42% of individuals with 2D and 3D replicates clustering together as "sisters"

TABLE 5 Correlation between variance covariance matrices (varcov), and paired tests (using 10,000 permutations) for differences in magnitude of sex-corrected shape variation between replicates: First versus second digitization for FULL-0 and 2D versus 3D in the other instances

Configuration	Statistic	r	Ratio[a]	Statistic	p	R^2 (%)
FULL-0	varcov	0.95				
	F ratio			0.302		0.4%
	VAR1		1.1	0.00015	.0549	
	VAR2		1.0	0.00239	.0463	
	VAR3		1.0	0.00186	.3737	
FULL-1	varcov	0.75				
	F ratio			3.604		4.9%
	VAR1		1.2	0.00044	**.0007**	
	VAR2		1.1	0.00715	**.0004**	
	VAR3		1.1	0.00736	.0244	
RED-4	varcov	0.81				
	F ratio		-	4.056		5.5%
	VAR1		1.3	0.00053	**.0002**	
	VAR2		1.1	0.00838	**.0002**	
	VAR3		1.1	0.01046	**.0071**	

[a]Second to first replicate for 2D digitizations and 3D to 2D for TTD data: thus, for instance, if >1 for FULL-1 or RED-4, that means that 3D shape has more variance than 2D.

TABLE 6 Summary statistics for covariances of sex-corrected shapes: Matrix correlation between individual and measurement error (ME) covariances (signed and excluding variances) and comparison of absolute covariances between individuals, ME and simulated ME with no covariance

Configuration	r indiv. vs. ME	Obs. vs. simulated	Summary statistics	Abs. Covariances	Ratios
FULL-0	0.25	Obs.: median	Individuals	0.00000314	31.4
		Obs.: median	2D digitizing error	0.00000010	1.3
		Simulated: median	Random error	0.00000008	
		Simulated: 95th percentile	Random error	0.00000008	
FULL-1	0.15	Obs.: median	Individuals	0.00000108	3.4
		Obs.: median	TTD error	0.00000032	1.3
		Simulated: median	Random error	0.00000024	
		Simulated: 95th percentile	Random error	0.00000025	
RED-4	0.16	Obs.: median	Individuals	0.00000141	3.5
		Obs.: median	TTD error	0.00000040	1.4
		Simulated: median	Random error	0.00000028	
		Simulated: 95th percentile	Random error	0.00000030	

Note: Ratios of medians (individual vs. ME and ME vs. "random ME") are shown in the last column.

(a) full config. 2D digit. error vs individual

(b) full config. TTD error vs individual

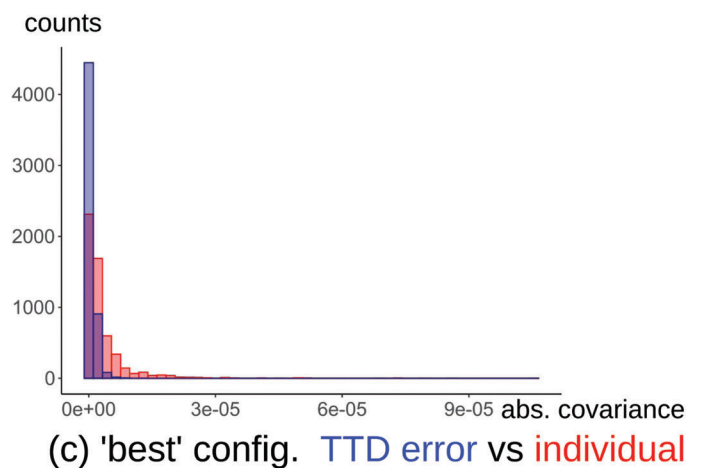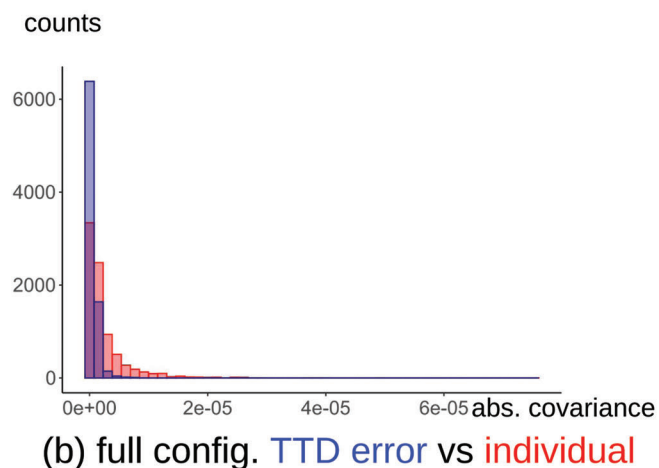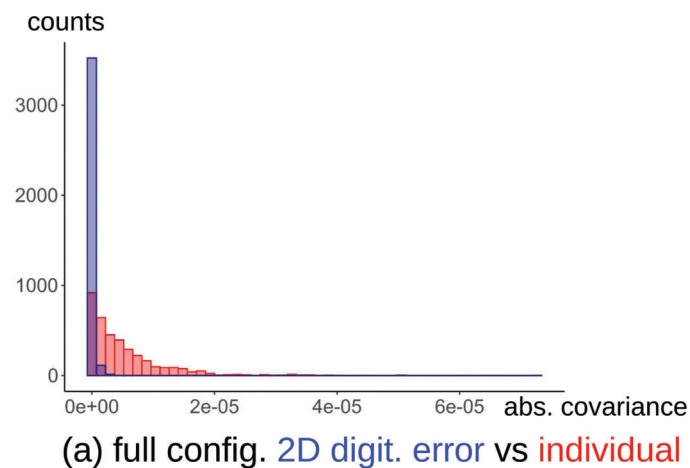**(c)** 'best' config. TTD error vs individual

FIGURE 5 Distribution of absolute covariances for the sex-corrected individuals (red) and the ME residuals (blue): (a) FULL-0; (b) FULL-1; and (c) RED-4
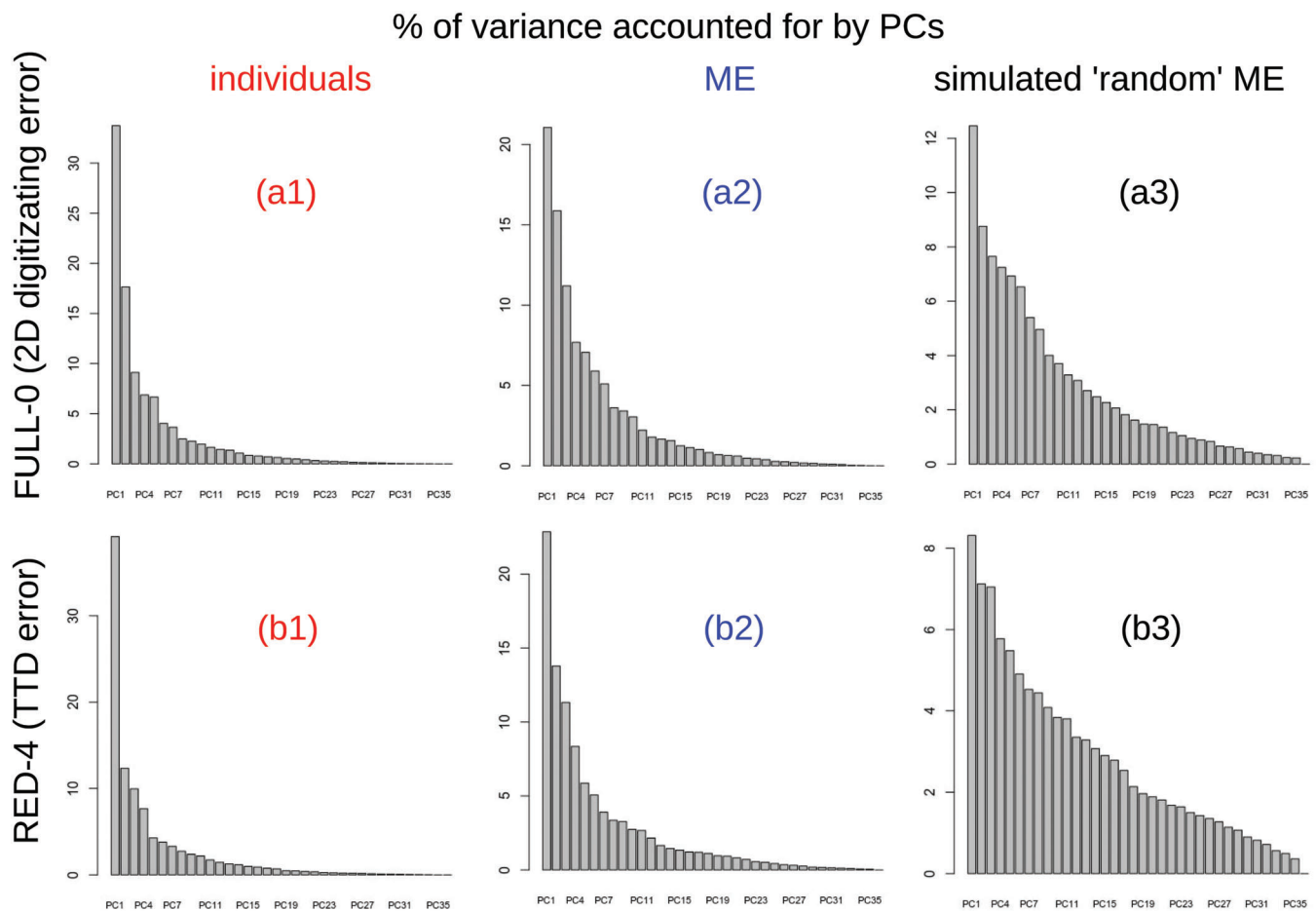
FIGURE 6 Percentages of sex-corrected shape variance accounted for by PCs in FULL-0 (a) and RED-4 (b): (a1–b1) individuals; (a2) 2D digitizing error; (a3) example of simulated "random error" with the same variance as a2 but no covariance; (b2) TTD error; (b3) simulated "random error" with the same variance as b2 but no covariance

Cardini, 2019). Thus, as evident in Figures 6a3,b3, variance accounted for by simulated random noise decreases fairly smoothly as one moves from the first to higher order PCs. In contrast, variances accounted for by either digitizing (Figure 6a2) or TTD error (Figure 6b2) are somewhat in between the strong pattern of individual shapes and the gradual pattern of random data. This is suggested by PC1 of ME explaining some 50% less variance than PC1 of individuals but about twice the variance explained by PC1 of simulated random noise. It should also be noted that the difference between observed ME and random noise is particularly pronounced for TTD data.

Table 6 and Figure 7 show the overall results of the simulations, which confirm the impression from Figures 5 and 6. ME has a median absolute covariance above the 95th percentile of the medians in the 100 simulated datasets and is, therefore, significantly larger (ca. 30–40%) than expected in random variables with the same variance but no real covariance (except for that due to sampling error and the superimposition). The median absolute covariance of ME is, however, much smaller than for individuals. More precisely, the median absolute covariance of individuals is more than 30 times larger than that of 2D digitizing error (FULL-0) and about 3–4 times larger than the TTD error (FULL-1 and RED-4). Overall, these results suggest that ME is not random and has structure, but weak when compared to covariation in relation to differences among individuals.

3.4 | Parallel tests of taxonomic differences

Analyses restricted to western patas and eastern patas (after controlling for sex) confirm the very high congruence of size data, with 2D and 3D suggesting virtually identical patterns of differences (Figure 8): eastern patas have, on average, larger crania, but the range of sizes in the two taxa overlaps extensively. As anticipated, all further analyses were done only on sex-corrected shape, whose larger ME makes crucial the question about whether 2D and 3D shape results are similar.
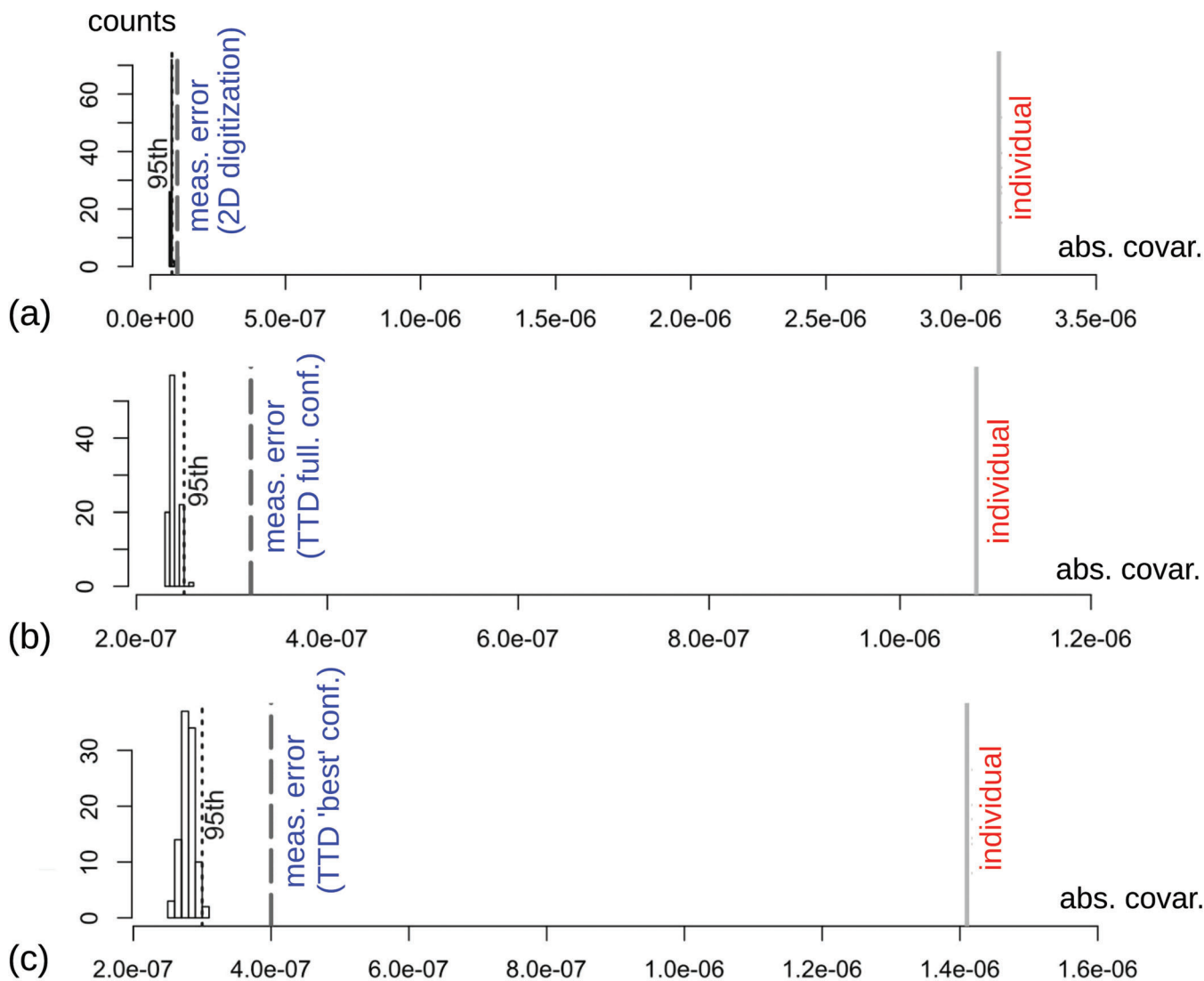
FIGURE 7 Median of absolute covariances of sex-corrected shape coordinates for individuals (light gray), ME (dark gray long dash), and "random error" (histogram of medians from 100 runs of simulation with dotted line marking their 95th percentile): (a) FULL-0; (b) FULL-1; (c) RED-4

3.4.1 | Differences in mean shapes between taxa

Tests of group differences in shape using the FULL configuration, as well as RED-4, confirm the strong effect of sex (accounting for ca. 40% of variance) and suggest negligible differences between western patas and eastern patas (Table 7). Mean differences are never significant and account for less than 3% of shape variance. The interaction between sex and taxon is also small (R^2 = ca. 4%) and does not reach significance. This supports the appropriateness of controlling for sex, as patterns do not differ statistically between the two groups. As for other tests, however, nonsignificance may be due to the low power imposed by small samples.

Cross-validated classification accuracy ranges, on average, from 55% to 65%. This is only slightly better than chance (median baseline = 48–52%), and never above the 95th percentile of 66–72% required for significance. Thus, the XbgPCA classification also indicates negligible differences in shape regardless of the configuration and type of data.

Indeed, the congruence between 2D and 3D data is almost perfect in terms of conclusions about significance but also, and more importantly in terms of estimates of the magnitude of the differences, with very similar R^2s and hit rates. In fact, differences in results of tests for differences between taxa are slightly larger between the first and second digitization of 2D data than between 2D and 3D data. For instance, for the effect of taxon, the FULL-0
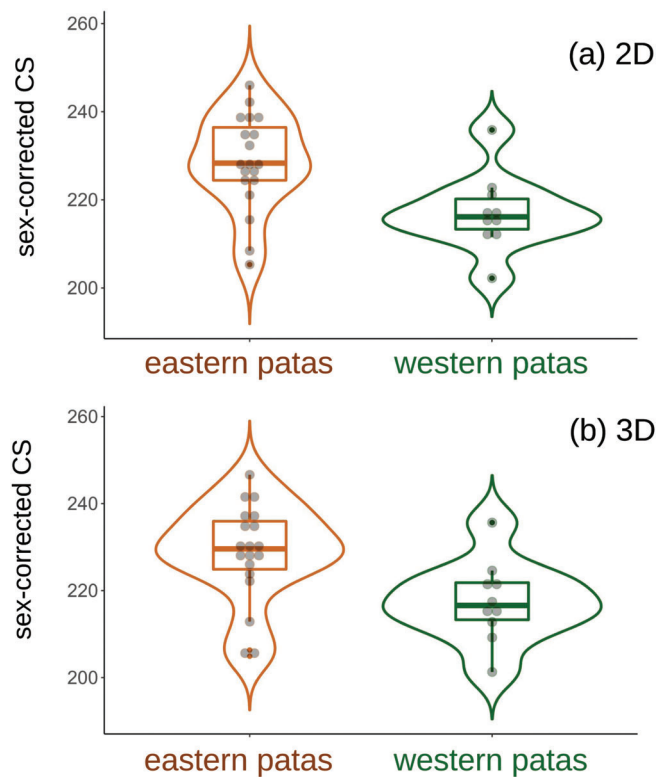
FIGURE 8 Centroid size of RED-4 estimated in 2D (a) and 3D (b) for sex-corrected eastern patas and western patas. The shape of the violin plots, the distribution of points in the jitter plots, and the similar box plots all indicate an excellent congruence of the two types of data

R^2 difference between 2D replicates is 0.4%, whereas for 3D versus 2D in FULL-1 and RED-4 it is, respectively, 0.2% and <0.1%. Similarly, with hit rates, the first and second 2D digitization show a difference of 7% (with the first digitization having higher classification accuracy), whereas the corresponding differences between 3D and 2D in FULL-1 and RED-4 are, respectively, 4% and 0% (with 3D shapes being marginally more accurate in group prediction only in FULL-1).

Scatterplots of ordinations summarizing RED-4 shape data also confirm the results of the statistical tests with an almost complete overlap of western patas and eastern patas (Figure 9). PCAs show a small separation of their respective means both in 3D and 2D data (Figure 9a1,b1) and slight differences in patterns of variation between the two taxa. These are suggested by the orientation and elongation of the confidence envelopes which, however, are not always congruent between 2D and 3D. In contrast, XbgPCA scatterplots (Figures 9a2,b2) show almost identical patterns both in 2D and 3D, with very close means and overlapping confidence envelopes consistently vertically stretched

(i.e., in the main direction of nonbetween group variance). Ordinations (not shown) for FULL-1 suggest similar conclusions of strong congruence with mostly minor differences between 2D and 3D data.

Shape diagrams of RED-4 mean shapes of western patas and eastern patas in 2D and 3D (Figure 10) must be magnified 10 times to visualize small differences. We do not describe these nonsignificant differences, as they are tiny. In contrast, it is interesting to observe that, even in spite of the huge magnification, 3D and 2D produce an almost perfect match in shape changes in all anatomical regions of the ventral cranium except the basisphenoid-basioccipital region. Here, 2D shows a widening of the bone and slight enlargement of the occipital foramen in western patas compared to eastern patas, whereas 3D suggests the opposite. Without magnification, however, the 2D and 3D patterns look almost identical, both showing very minor differences between taxa.

3.4.2 | Differences between taxa in magnitude of sex-corrected shape variance

None of the tests (first or second 2D digitization or 2D vs. 3D) show appreciable differences in shape variance between western patas and eastern patas (Table 8), although the latter, with its larger *N*, consistently has slightly larger values. 2D, however, tends to moderately overestimate this small difference showing larger R^2s (ranging from ca. 2% to 6%) and variance ratios. This suggests that eastern patas vary approximately 10% more than western patas. In contrast, 3D shapes have smaller R^2s (ca. 1%), with variance in eastern patas only 3% larger, on average, than in western patas. Thus, although the agreement between 2D and 3D results is slightly less precise than between the first and second 2D digitization (unlike what we found with group mean differences in Section 3.4.1), the bias in 2D is small and does not affect the conclusions of the tests.

3.4.3 | Static allometric trajectories of sex-corrected shapes in western patas and eastern patas

Results of tests of taxonomic differences in static allometry are also highly congruent between 2D and 3D data (Table 9). None of the factors in the multivariate ANCOVAs is significant in any of the datasets except for centroid size, which is always highly significant ($p < .01$) and accounts for 10–11% of variance in the total configuration (FULL-0 and FULL-1), and slightly more (13%) in RED-4. The effect of taxon, as well as the effect of the interaction of taxon and centroid size,

TABLE 7 Multivariate analysis of variances (ANOVAs) for shape differences in 3D and 2D, and leave-one out cross-validated classifications using sex-corrected shapes

| 2D or 3D | Type of shape data | ANOVA | | | | | | | bgPCA | Random baseline | |
		Factor	df	SS	MS	F	R²	p	Hit rate	Median	95th
2D	FULL-0 1st digitization	Sex	1	0.03108	0.031081	18.51	39.4%	*.0001*			
		Taxon	1	0.00222	0.002215	1.32	2.8%	.2337	62.1%	52.2%	69.2%
		Sex by taxon	1	0.00355	0.003553	2.12	4.5%	.0809			
		Residuals	25	0.04198	0.001679		53.3%				
		Total	28	0.07882							
	2nd digitization	Sex	1	0.03229	0.032290	21.02	42.9%	*.0001*			
		Taxon	1	0.00184	0.001839	1.20	2.4%	.2720	55.2%	48.3%	69.2%
		Sex by taxon	1	0.00278	0.002775	1.81	3.7%	.1248			
		Residuals	25	0.03840	0.001536		51.0%				
		Total	28	0.07531							
	FULL-1[a]	Sex	1	0.03101	0.031005	20.76	42.3%	*.0001*			
		Taxon	1	0.00193	0.001925	1.29	2.6%	.2365	62.1%	55.2%	69.0%
		Sex by taxon	1	0.00302	0.003016	2.02	4.1%	.0981			
		Residuals	25	0.03734	0.001494		51.0%				
		Total	28	0.07329							
	RED-4[b]	Sex	1	0.02742	0.027423	19.01	40.9%	**0.0001**			
		Taxon	1	0.00153	0.001533	1.06	2.3%	.3305	62.1%	48.3%	65.7%
		Sex by taxon	1	0.00203	0.002026	1.40	3.0%	.2046			
		Residuals	25	0.03607	0.001443		53.8%				
		Total	28	0.06705							
3D	FULL-1[a]	Sex	1	0.03602	0.036022	20.34	42.2%	*.0001*			
		Taxon	1	0.00201	0.002010	1.14	2.4%	.2745	65.5%	51.7%	69.0%
		Sex by taxon	1	0.00302	0.003021	1.71	3.5%	.1283			
		Residuals	25	0.04427	0.001771		51.9%				
		Total	28	0.08532							
	RED-4[b]	Sex	1	0.03334	0.033341	18.98	40.6%	**0.0001**			
		Taxon	1	0.00189	0.001889	1.08	2.3%	.3131	62.1%	51.7%	72.4%
		Sex by taxon	1	0.00289	0.002886	1.64	3.5%	.1391			
		Residuals	25	0.04392	0.001757		53.5%				
		Total	28	0.08204							

[a]2D versus 3D matrix correlation for sex-corrected Euclidean shape distances $r = 0.58$.
[b]2D versus 3D matrix correlation for sex-corrected Euclidean shape distances $r = 0.54$.

which tests whether allometries differ between western patas and eastern patas, are not significant. Indeed, they account for very small amounts of shape variance (2–4%). Nevertheless, in the highly dimensional shape spaces, the angles formed by the allometric trajectories of the two patas taxa are large (ca. 49–50° on average). Angles are very similar in 3D and 2D being, respectively, 53–55° in FULL-1 and 48–50° in RED-4. In fact, the discrepancy in estimates of the angles of allometric vectors is larger in the comparison of the first

and second 2D digitization, with a difference of 5° (53° vs. 58°), compared to just 2° in the TTD datasets.

4 | DISCUSSION

4.1 | Aim and context

ME in GMM has received more attention in the last decade, as shown by Fruciano (2016) in his detailed
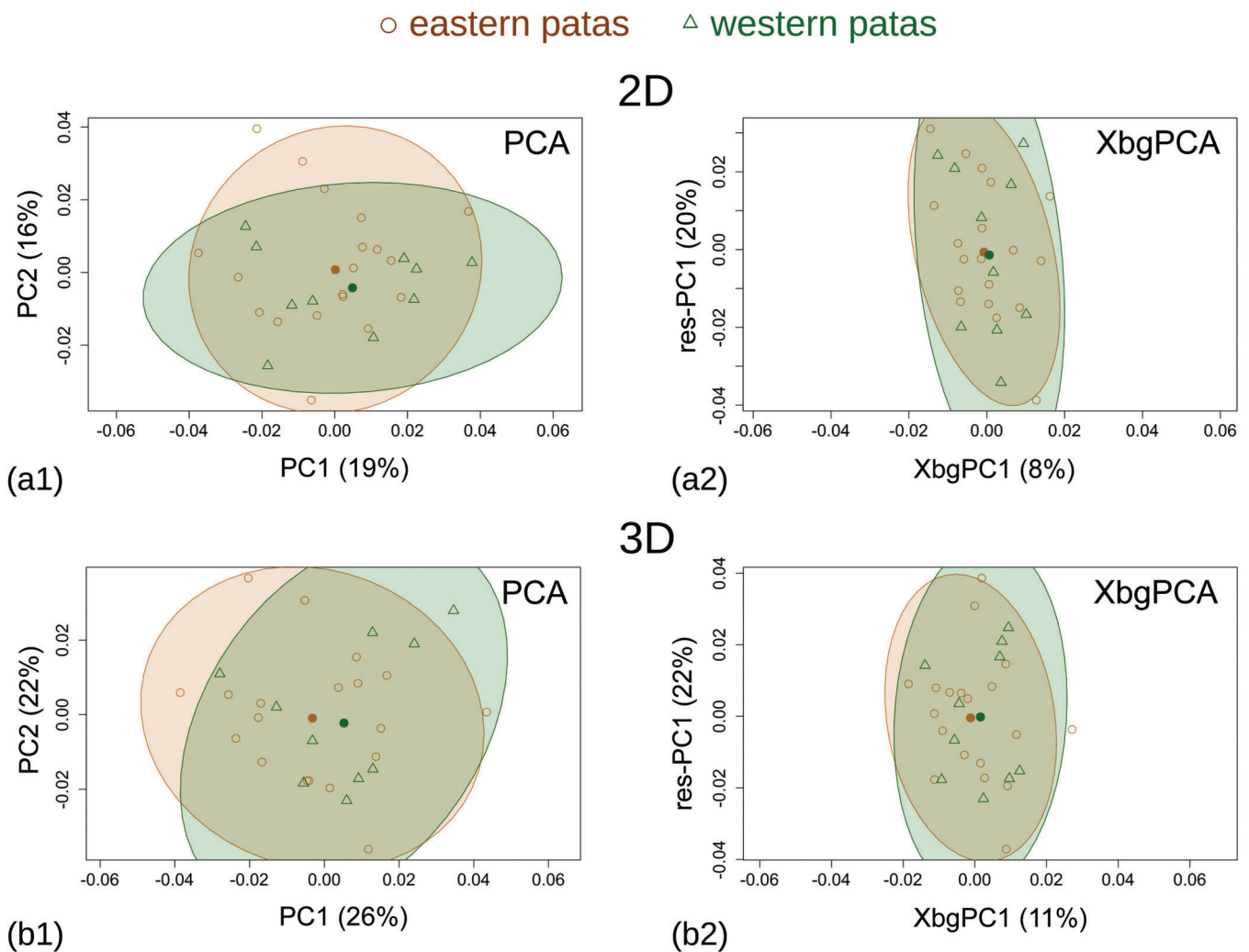
FIGURE 9 Scatterplots summarizing RED-4 sex-corrected shape variance (percentages of total in parentheses) of eastern patas and western patas ventral crania with 2D and 3D data in separate shape spaces (95% confidence envelopes are shown as well as mean shapes [filled circles] for the two groups). In the PCA (a1–b1), 2D and 3D both suggest a large overlap between eastern patas and western patas, although with a more elongated scatter for western patas, especially in 2D. The XbgPCAs (a2–b2, with res-PC1 representing the main axis of nonbetween group variance) confirm, regardless of the type of data, the lack of appreciable mean differences and the almost complete overlap of the two groups
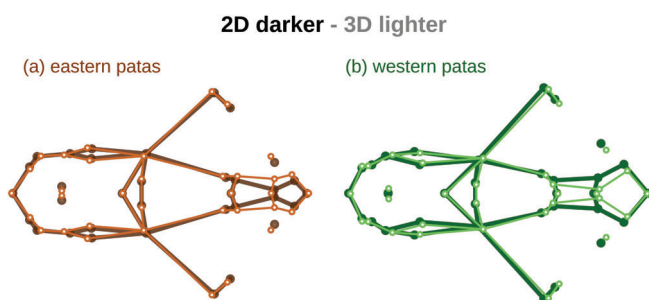

FIGURE 10 RED-4 average sex-corrected shapes of (a) eastern patas and (b) western patas magnified 10×. 3D means (lighter and thinner lines) were manually superimposed on 2D means (darker and thicker lines) to emphasize differences. Despite the huge magnification, they overlap almost perfectly except in the basisphenoid-basioccipital region

review, but also as summarized, in the specific context of TTD errors, by Cardini and Chiapelli (2020). The renewed interest in ME is welcome, as the field has encountered increasing success over the years (Adams et al., 2013), but most of its focus has been on new methods and innovative applications (Cardini & Loy, 2013). Much less attention has been devoted to the centrality of accurate data (Cardini, 2020a) and other less "fashionable" topics, including ME. Arnqvist and Martensson (1998) wrote their important and highly cited contribution on ME in GMM almost 20 years before Fruciano's (2016) much needed update. Exactly 20 years have passed since Roth's (1993) early, but largely ignored, warning on the potential problems of 2D studies and the publication of a series of papers

TABLE 8 Tests for differences in magnitude of sex-corrected shape variation between western and eastern patas analyzed separately for the first and second replicate (2D digitizations) or 3D and 2D data

Configuration	Test statistics	1st replicate or 3D				2nd replicate or 2D			
		Ratio[a]	Statistic	p	R^2 (%)	Ratio	Statistic	p	R^2 (%)
FULL-O	F ratio		1.014	.3265	3.6%		0.444	.5119	1.6%
	VAR1	0.9	0.00026	.4440		0.9	0.00016	.6428	
	VAR2	0.9	0.00419	.4628		1.0	0.00231	.6877	
	VAR3	0.9	0.00799	.3860		0.9	0.00783	.4888	
FULL-1	F ratio		0.234	.6365	0.9%		0.667	.4278	2.4%
	VAR1	1.0	0.00006	.8142		0.9	0.00018	.5668	
	VAR2	1.0	0.00074	.8533		0.9	0.00293	.5989	
	VAR3	1.0	0.00291	.6749		0.9	0.00593	.5071	
RED-4	F ratio		0.305	.5856	1.1%		1.652	.2094	5.8%
	VAR1	1.0	0.00007	.7896		0.8	0.00024	.3593	
	VAR2	1.0	0.00107	.7962		0.9	0.00451	.3423	
	VAR3	1.0	0.00150	.7948		0.9	0.00689	.3150	

[a]Ratio between VAR in western and eastern patas; ratios <1 indicate larger variance in eastern patas.

on TTD errors (see Cardini & Chiapelli, 2020 for references) initiated by Álvarez and Perez (2013) and Cardini (2014).

ME is especially important for a careful assessment of small taxonomic differences among closely related taxa. 2D photographs, as we explained in the Introduction, are a very convenient source of information for taxonomic studies using morphological data. They are relatively easy to acquire and low cost. However, if measurements on photographs provide an inaccurate representation of the real 3D structures, results can be misleading. The validity of 2D data is generally taken for granted and hardly mentioned in most GMM analyses using photographs. Yet, this implicit assumption is crucial, should be discussed and, whenever possible, tested at least with a small sample in order to explore the appropriateness of the TTD approximation in relation to the specific study question.

As the main research on TTD errors has been concisely presented by the recent paper of Cardini and Chiapelli (2020), we refer readers to that article for an overview of the main findings and approaches. Here we focus our discussion on a comparison with two other studies (Cardini, 2014; Cardini & Chiapelli, 2020) that, as in this paper, consistently used all three main methods to explore TTD congruence. Thus, we discuss findings using the approach of Cardini (2014), which tests TTD error within a common 2D–3D shape space, as well as findings from correlational analyses and the comparison of results from tests run in parallel in 2D and 3D. These previous studies are not only more directly comparable to our study of patas, but also provide a complementary point of view on cranial TTD approximation in different groups of mammals. Cardini (2014) used marmots as a case study. Marmots are are rodents with an adult body mass of about 3–7 kg (Armitage, 1999) and crania about 8 cm in length (Cardini, unpublished). Cardini and Chiapelli (2020), in contrast, analyzed much larger animals, the living equids, which are typical representatives of the terrestrial megafauna, with a body mass at least 25 times (Clauss et al., 2009) that of the largest marmot and crania 50 cm or more in length. In terms of size, patas crania are somewhat in between. Although much smaller than equids, patas have an average adult body mass (ca. 7 kg for female and 12 kg for male *E. p. patas*; Isbell, 2013) about twice that of marmots, and cranial lengths of 8–11 cm.

Patas are, to our knowledge, the first example of an in depth study of TTD errors in primates. Thus, we work with representatives of a different order of placentals (Primates), but within the same evolutionary and taxonomic context as Cardini (2014) and Cardini and Chiapelli (2020), who focused on the relatively small variation typical of microevolutionary studies of adults within the same species or, at the boundary between micro- and macroevolution, within the same genus. The analyses we performed on patas are, nevertheless, more extensive, as they include differences in the magnitude and structure of shape variance–covariance, as well as an exploratory examination of the covariance structure of ME. In the next sections, starting with size followed by

TABLE 9 Multivariate ANCOVAs for allometry using sex-corrected shape and size in 3D and 2D

2D or 3D	Type of shape data	ANCOVA Factor	df	SS	MS	f	R^2	p
2D	FULL-0 1st digitization	Taxon	1	0.00202	0.002019	1.26	4.2%	.2302
		CS	1	0.00469	0.004687	2.93	9.8%	**.0073**
		Taxon by CS	1	0.00107	0.001068	0.67	2.2%	.7470
		Residuals	25	0.03997	0.001599	0.84		
		Total	28	0.04774	1.000000			
	2nd digitization	Taxon	1	0.00168	0.001676	1.16	3.9%	.2820
		CS	1	0.00448	0.004479	3.11	10.4%	**.0058**
		Taxon by CS	1	0.00088	0.000883	0.61	2.1%	.7924
		Residuals	25	0.03598	0.001439	0.84		
		Total	28	0.04302				
	FULL-1	Taxon	1	0.00176	0.001755	1.25	4.2%	0.2472
		CS	1	0.00449	0.004487	3.19	10.6%	**.0046**
		Taxon by CS	1	0.00087	0.000874	0.62	2.1%	.7783
		Residuals	25	0.03517	0.001407	0.83		
		Total	28	0.04229				
	RED-4	Taxon	1	0.00140	0.001397	1.08	3.5%	0.3484
		CS	1	0.00515	0.005145	3.99	13.0%	**.0002**
		Taxon by CS	1	0.00082	0.000819	0.63	2.1%	.7838
		Residuals	25	0.03227	0.001291	0.81		
		Total	28	0.03963				
3D	FULL-1	Taxon	1	0.00183	0.001832	1.13	3.7%	.2978
		CS	1	0.00538	0.005382	3.31	10.9%	**.0001**
		Taxon by CS	1	0.00149	0.001493	0.92	3.0%	.5369
		Residuals	25	0.04059	0.001624	0.82		
		Total	28	0.04930				
	RED-4	Taxon	1	0.00172	0.001722	1.08	3.5%	.3487
		CS	1	0.00607	0.006067	3.82	12.5%	**.0001**
		Taxon by CS	1	0.00123	0.001227	0.77	2.5%	.7188
		Residuals	25	0.03968	0.001587	0.81		
		Total	28	0.04869	1.000000			

the main analysis on shape, we summarize the most important findings and compare them with previous research.

4.2 | Centroid size: Excellent TTD approximation, and the importance of coplanar landmarks

ANOVAs, correlations, and summary plots all confirmed that 2D data provide estimates of centroid size which are almost perfectly congruent with those from 3D measurements. This is in agreement with Cardini ([2014](#)) and Cardini and Chiapelli ([2020](#)), hence our focus on cranial shape. Indeed, 2D centroid size is accurate and shows no strong bias in patas ventral crania. For instance, using all landmarks (FULL-1), the average deviation in western patas and eastern patas is within ±1–2 mm of the corresponding 3D estimates, which in relative terms translates to an inaccuracy of less than 1%. For comparison, the same differences in mean centroid size between the first and second 2D replicates (FULL-0) are, on average, about 0.5 mm, but that is only 2D digitization error on the same

photographs and would probably be about twice as large (see the relative discussion in Section 2) if specimens had been repositioned and rephotographed. Thus, TTD error in patas cranial size is only slightly larger than expected in replicates of 2D landmarks and smaller than TTD inaccuracies in marmots (Cardini, 2014) and equids (Cardini & Chiapelli, 2020), in which 2D centroid size was precise but on average tended to slightly (ca. 2% or less) over-estimate (marmot lateral and ventral cranial views) or under-estimate (marmot hemimandibles and equid ventral crania) 3D centroid size.

Two-dimensional centroid size is, in fact, expected to be biased and typically underestimate the size of a 3D structure because distances between landmarks and their centroid are smaller when variation in depth cannot be measured. Depending on its position, however, relative to the landmarks, the scale factor in a photograph may consistently bias measurements in two directions (downward, if placed higher than most landmarks, and upward, if lower). For instance, the position of the scale factor below the landmarks was the likely explanation for centroid size overestimates in marmot crania (Cardini, 2014). Measuring a long interlandmark distance with a caliper, such as condylobasal length, directly in 3D, and using this measure to rescale landmarks from 2D photographs, may avoid issues with the scale factor in the photograph and contribute to reducing error in estimates of size. For accuracy it is even more important that landmarks are mostly coplanar.

In the ventral crania of patas, only one landmark (inion, Figures 2 and 3) lies well outside the main plane defined by the configuration. It is not always easy to know precisely which landmarks might be less coplanar than others. Cardini (2014) argued that the zygomatic arches may often be problematic, but that is specific to the anatomy of the study taxon and the orientation of the cranium. In side view, for example, it is very common to have landmarks on the midplane, as well as around teeth, orbits, and zygoma (Alhajeri, 2018; Bohórquez-Herrera et al., 2017; Borges et al., 2017; Cardini et al., 2005; Chemisquy, 2015; Chevret et al., 2020; D'Anatro & Lessa, 2006; dos Reis et al., 2002; Evin et al., 2008; Fornel et al., 2010; Lalis et al., 2009; Loveless et al., 2016; Marcy et al., 2016; Milenvić et al., 2010; Myers et al., 1996; Panchetti et al., 2008; Pandolfi et al., 2020; Samuels, 2009; Scalici et al., 2018; Yazdi, 2017; Yazdi et al., 2012). These inevitably span multiple depths and may strongly distort estimates of size and shape in closely related species. Thus, coplanarity seems crucial in studies of small differences and, together with rescaling the raw coordinates using a 3D distance, likely explains why the inaccuracy in centroid size estimates is so small in our dataset and the error seems random, with no consistent bias. This observation strengthens the conclusion of previous research on marmots and equids that, at least in

crania of mammals of medium or large size, centroid size can be accurately measured using photographs as long as the photographic setting is reasonably standardized, the landmarks are mostly coplanar, and the scaling factor is accurate.

4.3 | Magnitude of TTD error in shape

Unlike for size, the magnitude of TTD error for shape is large. Individual variation within sex-corrected patas is only 4–5 times larger than differences between 2D and 3D, which is similar to findings in marmot crania (Cardini, 2014) but slightly more than in equids (Cardini & Chiapelli, 2020) (with individual variance about four - marmots - and three - equids - times TTD error variance). Only in the relatively flat marmot hemimandibles were individual differences much larger (approximately nine times) than TTD error (Cardini, 2014) but, in that dataset, individual variance was slightly inflated by the inclusion of a few young animals in a sample of mostly adult animals. Thus, overall, it seems that TTD error, especially within fairly homogeneous samples of adult mammal crania, accounts for a large proportion of shape variance within a species or, in the case of patas, between putative subspecies.

Individual variation was, nevertheless, significantly larger than TTD error in all three studies, but significance on its own does not mean that ME is negligible. The ANOVA test simply rejects the null hypothesis that differences among individuals are as large as differences between 2D and 3D, but cannot rule out the possibility that the latter are large enough to make 2D results inaccurate. In fact, if individual variation is only about 3–5 times larger than TTD error, then the inaccuracy in the approximation of 3D shapes in 2D has an effect size about as large as interspecific differences within a genus of mammal. For instance, individual differences within a species of marmot (Cardini, 2014) or equid (Cardini & Chiapelli, 2020) are about 2.5–5.5 times larger than interspecific variation, which is the same range as when individual variation is compared with TTD error.

The relatively modest congruence of 2D and 3D shapes is evident also by the moderate correlations between corresponding matrices of Procrustes shape distances. In the ventral crania of equids (Cardini & Chiapelli, 2020), r ranged from about 0.5 to 0.6 (respectively, within plains zebra or *Equus* regardless of species), while in marmots (Cardini, 2014) it ranged from 0.5 to 0.7 for crania and up to 0.84 for hemi-mandibles. If we only consider cranial data, which are more directly comparable across taxa, these correlations are only slightly

smaller than in patas ($r \approx 0.7$–0.8). The difference is unlikely to be a consequence of the smaller size of the patas sample. Cardini and Chiapelli (2020) showed that small random subsamples of either plains zebras or the total *Equus* sample produce a larger range of values of matrix correlations (i.e., lower precision) but, on average, estimates are very similar to (actually, very slightly smaller than) those observed including all specimens. Indeed, in small samples, unlike R^2 that is biased upward (Cramer, 1987), r tends to be underestimated, especially when in the range of approximately 0.5–0.8 (Zimmerman et al., 2003). Thus, at least compared to the homogeneous intraspecific sample of plains zebras, the larger r of sex-corrected patas ventral cranial shapes might be explained by either a really smaller TTD error or by potentially larger differences among individuals (or by a mix of these two effects). That our small sample of patas might span a range of differences larger than found within a single species is consistent with the recent revision of the genus, that suggests that *Erythrocebus* is a species complex (De Jong & Butynski, 2020a,b, 2021; De Jong et al., 2020; Gippoliti & Rylands, 2020). A comparatively larger variability among specimens in the sample is also a likely reason why individual differences in patas are 4–5 times larger than TTD error, but only three times larger in plains zebras and other equids.

Configurations with a smaller number of landmarks may have lower correlations of 2D and 3D Procrustes shape distances. This was seen in equids, although the effect was small and almost negligible (Cardini & Chiapelli, 2020). In fact, a configuration of just about two dozen landmarks, as in Cardini and Chiapelli (2020) and in this study of patas, does not allow a proper assessment of whether having more or fewer landmarks has an effect on the goodness of the TTD approximation. Indeed, fewer landmarks do not automatically lower 2D accuracy, as we have shown that a careful selection of a subset of landmarks can slightly improve the correspondence between 2D and 3D shapes of patas.

On the other hand, the choice of landmarks, and therefore their number is, strictly functional to the specific study question (Cardini, 2020a, 2020b; Klingenberg, 2008; Oxnard & O'Higgins, 2009). If the aim is individual identification, as in forensics and some other disciplines, a larger set of points might be a good choice for capturing the often minute anatomical details that differentiate each individual. However, for taxonomic comparisons and many other applications in evolutionary biology or ecology, where the purpose is to accurately describe variability within and among taxa, a better choice could be a smaller but carefully designed set of points which capture the main differences in relative proportions of a structure, leaving out

noisy details of dubious relevance for population biology. This type of trade-off might explain why Cardini (2014) found that ventral crania, with their larger set of landmarks, had a larger proportion of individuals clustering as sister replicates in phenograms of shape distances (ca. 60% vs. ca. 40–55% in lateral and dorsal views of the cranium), despite a lower correspondence between 2D and 3D distances (matrix $r \approx 0.5$ vs. ca. 0.6–0.8 in other views), as well as between variance–covariance matrices (matrix $r \approx 0.5$ vs. ca. 0.7 in other views). On the effect of the number and type of landmarks on 2D accuracy, there is not enough evidence at present to make any recommendation. We speculate that sometimes an increase in precision at the level of the individual comes at the cost of a reduction in overall accuracy in the quantification of inter-individual similarity relationships, but this will have to be assessed in future research.

Cardini (2014) and Cardini and Chiapelli (2020) did not test differences in variance between 2D and 3D data in the common shape space. Therefore, we cannot say if any difference they reported was significant or not, but in marmots, as in patas, 3D shape variance was consistently larger than 2D variance. One exception, however, is that 2D scans of hemi-mandibles showed slightly more variance in 2D. A larger (ca. 8–27%) 2D shape variance was also found in equids. This is unexpected, because 2D shape has, in fact, zero variation on the Z axis (the third "fake" coordinate added to superimpose data in a common shape space). Variance should, therefore, be larger in 3D, as in patas and most of the marmot datasets (e.g., Cardini, 2014). So, why was 2D shape variance larger in ventral crania of equids, as well as in the scanned marmot hemi-mandibles? It could be because of difficulties of standardizing the orientation of a specimen in large animals such as equids, or relate to the type of instrument used for acquiring images. For instance, hemi-mandibles laid on a flat-bed scanner are less easy to be consistently and precisely oriented in the same way, because they lie on the relatively irregular surface of the lingual side. Thus, the inaccuracies in replicating the orientation of different specimens adds some extra variance to the data. This also explained why scans of marmot hemi-mandibles had a larger ME than photographs (Cardini, 2014). In the case of equids, it is also possible that, besides issues with standardizing the orientation, photographic distortions have contributed to variance inflation. This seems likely, because the distance between the camera and the specimen was relatively short and a wide-angle zoom had to be used to photograph these large crania (Mullin & Taylor, 2002). Marmot and patas crania not only are easier to position but also, being much smaller,

allow the operator to increase the distance between the specimen and the camera to reduce distortions in the photographs (Cardini & Tongiorgi, 2003). Both inconsistencies in the orientation of the cranium and photographic distortions may have inflated 2D shape variance and contributed to making TTD error in equids (Cardini & Chiapelli, 2020) larger than in marmots (Cardini, 2014) and patas.

4.4 | Can we reconcile a modest TTD approximation with accurate 2D results?

Considering the large proportion of variance accounted for by TTD error in patas, and the low percentage ($\leq 50\%$) of individuals clustering as "sister" replicates in the phenograms, despite the moderately large 2D–3D correlations of Procrustes shape distances and variance–covariance matrices, a complete congruence between tests performed in parallel on 2D and 3D shapes is unexpected. Yet, our series of tests of "biological hypotheses" shows an almost perfect correspondence in results of 2D and 3D shape analyses. Even more surprisingly, this happens not only for factors with a large effect (such as sex differences or centroid size in the ANCOVAs), but also for those accounting for very small proportions of variance (e.g., taxonomic differences in means, variances, and allometric trajectories). There are a few exceptions where estimates in 2D deviate slightly more from the corresponding 3D estimates, but even in these instances the conclusions from the tests are the same.

Several of the factors we tested show a very small effect size and do not reach significance. Power is low in our small samples and robust answers require many more specimens, but this is of secondary interest in this methodological study on the degree of congruence between 2D and 3D results. In fact, having mostly tiny deviations between 2D and 3D estimates of small R^2s, as well as between estimates of classification accuracy using shape, is counter-intuitive. Like sampling error, ME is also likely to have a stronger impact on estimates of small differences. In our study, the variance explained by the factors being tested is often half the variance accounted for by TTD error, or even less than half. For instance, for RED-4, with a TTD error approximately equal to one-fifth of individual variance (Table 4), the variance explained by taxonomic differences was 2.3% both in 2D and 3D, which is just 1/23rd of individual variation (Table 7). RED-4 cross-validated hit rates (Table 7) were identical (62%), although with small differences in the 95th percentiles for the random baseline of classification accuracy (66% in 2D and 72% in 3D). Even this discrepancy is, however, small and probably partly relates to the modest number of randomizations.

With more randomizations, baselines percentages would likely converge toward more similar values.

That, in spite of the excellent congruence in the tests and summary plots (Figures 8 and 9), the visualization of mean shape differences between western patas and eastern patas suggest some differences between 2D and 3D (Figure 10) is understandable. The differences between these two taxa are small and had to be magnified 10 times to make them visible. Thus, it is probably more surprising that the visualization is almost identical for the rostrum, and only differs in the cranial base region. This difference, visible only after magnifying changes many times is, in fact, a good reminder not to over-interpret small and nonsignificant variation.

With highly congruent 2D and 3D results of tests, summary plots, and shape diagrams, an obvious question is: How might this happen when the magnitude of the TTD error is as large as, or larger than, the size of most of the effects being tested? This apparent contradiction is not new. Cardini and Chiapelli (2020) found the same in equids, where the error was even larger than in patas and yet results of the tests of "biological" hypotheses were perfectly congruent in terms of significance, magnitude, and even patterns in ordinations and shape diagrams. Thus, individuals are not in the same exact relative positions in the 2D shape space as in the 3D one, but the differences do not seem to lead to inaccurate inferences when samples are tested. Cardini and Chiapelli (2020) speculated that this paradox can be explained if TTD error adds a moderate amount of random noise to a strong pattern of "true" biological covariance. We show, however, that TTD error is not random. Its covariance structure is stronger than expected by chance because of sampling error (Figures 6 and 7) and seems also larger than with digitizing error (Figures 5 and 6).

A degree of covariance in TTD error is, in fact, more plausible than an expectation of random error. If large relative to the size of the effects being tested, random error would significantly disrupt the correspondence of 2D with 3D similarity relationships. Instead, if there are no huge photographic deformations and the orientation of the specimens is well standardized, the distortions in the photographs due to the flattening of the third dimension should be relatively similar in all individuals (at least within species or genera, where the amount of biological variation is typically small). For instance, landmarks on the zygomatic arch will be on a plane which is slightly below that of landmarks on the palate and this should happen in all individuals and samples. Thus, the TTD distortion is not purely random and does introduce a certain amount of covariance, which partly modifies the "true" pattern of covariation. Although not as small as for digitizing error (Table 5, Figures 6 and 7), TTD

covariances are, on average, more than three times smaller than individual covariance. Thus, after mean-centering, which removes the main bias between 2D and 3D shape due to the lack of information in the third dimension in 2D, the TTD error distorts the structure of variance and covariance, but the distortion is modest compared to the much stronger "true" variance–covariance. At least within genera of mammals, using cranial data, this inaccuracy seems to moderately alter the precise position of specimens in the shape space without having a strong impact on the general patterns of differences. It does not, therefore, alter the results of tests in 2D compared to the same tests in 3D.

Whether this interpretation is correct, and whether it can be generalized, remain open questions. Future studies will have to assess TTD error in other mammals and other organisms, and in other structures (e.g., postcranial bones or other views of the cranium), but also in relation to different hypotheses—for instance, tests of evolutionary trends or patterns of modularity and integration. For modularity/integration, where an accurate estimate of variance–covariance is crucial and Procrustes GMM already faces methodological issues (Cardini, 2019, 2020b), it will be particularly important to determine if a partial modification of the 3D pattern of covariance in 2D analyses adds a further layer of complexity and potential inaccuracy.

4.5 | Support for the accuracy of 2D Procrustes GMM using mostly coplanar landmarks, and considerations on the importance of "standardizing" photographs

Our research lends support to the observation that, at least when adult crania of closely related mammals are analyzed, results of 2D Procrustes GMM may be accurate despite TTD errors. This conclusion requires additional evidence but suggests that the "era" of 2D GMM is not yet over. 3D data are becoming easier to obtain and are more detailed and accurate whenever a structure is not flat. Collection of 3D data, however, requires instruments, such as 3D scanners, that are more expensive and typically slower and less portable than a digital camera. 3D photogrammetry, like 2D GMM, requires only a digital camera, but each specimen typically needs hundreds of photographs for accurate 3D reconstructions. This means longer time for collecting data and more computational power for processing them (Evin et al., 2016; Falkingham, 2012; Giacomini et al., 2019; Katz & Friess, 2014; Muñoz Muñoz et al., 2016). In addition, landmarking on 3D models is time-consuming and likely to take longer than digitizing the same landmarks on 2D photographs. Overall, it seems that 2D GMM offers a good alternative to 3D for a variety of applications. These include the measurement and testing of small taxonomic differences in morphology (this study, and Cardini & Chiapelli, 2020), but also the study of subtle covariation between genes and form (Navarro & Maga, 2016). Yet, accuracy cannot be taken for granted (e.g., Buser et al., 2018; Hedrick et al., 2019). Whether TTD error is negligible should be carefully considered before using 2D methods. This is even more crucial in studies of population, subspecies, or species differences, because the evidence they produce contributes to decisions on taxonomic status. These decisions, in turn, may influence the determination of conservation priorities (Mace, 2004).

Besides exploring the goodness of the TTD approximation in a preliminary study using, for instance, the truss method (Carpenter et al., 1996; Claude, 2008) to rapidly obtain low-cost 3D landmarks in a representative subsample (Cardini & Chiapelli, 2020), researchers must carefully plan how to obtain enough relevant information for the specific aim of their study, while minimizing TTD errors. The number, type, and position of landmarks (or semi-landmarks) are important, but this is not specific to 2D GMM (Cardini, 2020b). For 2D GMM, however, using fewer quasi-coplanar landmarks, and excluding those off the main plane of other landmarks, may be a useful expedient to reduce inaccuracies. Yet, one might want to first try including fewer coplanar landmarks, check the congruence with 3D in a subsample and, based on this, decide whether to include or exclude the landmarks which are off the main plane. In patas, for instance, we found a small but measurable improvement in the TTD approximation by excluding inion. There might be cases, however, when the exclusion of some landmarks is more problematic, as they may be in regions with crucial anatomical information. This could be one reason why 2D data on hystricognath (porcupine) hemi-mandibles, with a lateral flaring that makes them highly three-dimensional, have larger inaccuracies in 2D than those of the relatively flat sciurognath (squirrel) hemi-mandibles (Álvarez & Perez, 2013; Cardini, 2014; Cardini & Chiapelli, 2020).

The design of a highly standardized protocol for photographing specimens is fundamental for 2D GMM. Although our 2D data produced results largely congruent with 3D despite the poor standardization of the photographs, it is better to minimize potential sources of inaccuracy that can be easily controlled, such as the photographic settings. The orientation of the structure, or more precisely of the plane where most landmarks lie, must be kept as consistent as possible across all individuals. This requires not only determining in advance what landmarks to employ, but also how to place specimens

in the most appropriate position. Cardini and Tongiorgi (2003), for instance, built a small table whose inclination can be adjusted in order to check, using a spirit-level, that the lens and the specimen plane are parallel. This table has a frame of millimeter paper around its margins. The frame, thus, provides a scaling factor, but also helps to verify that the photograph shows no barrel-shaped deformations around the main central area. Marmot hemi-mandibles were placed horizontally on their buccal side on the small table, so that they always leaned the same way. This accurate placement may require the use of small pieces of plasticine to support the study structure. In the case of marmot hemi-mandibles, all but one landmark were on their outline in side-view and all were within a depth of <1 cm. This made it easier to standardize orientation without the use of plasticine. Finally, Cardini and Tongiorgi (2003) locked the camera on a portable copy stand with a high quality 180 mm APO lens and kept it as distant as possible (ca. 1 m) from the hemi-mandible in order to minimize photographic deformations. To keep the scale of photographic reproduction constant, they slightly adjusted, for each specimen, the height of the camera. Differences in height were very small (the lens aperture was kept at a minimum, and thus the width of field was just enough for the hemi-mandible to be in focus). This standardizes even more the distance of the camera to the chosen central point of focus (in this case, the proximal end of the diastema). They also worked with similar light settings in all photographs, and used diffuse lights to reduce shadows, since shadows can make some anatomical details harder to see.

Unlike the now obsolete analogical camera of Cardini and Tongiorgi (2003), modern high resolution digital cameras, held on a tripod and with a high-quality lens, simplify this protocol. Standardizing the orientation of crania is, however, less simple than with mandibles. For ventral views, one could replace the lid of a box with a frame of millimeter paper and have a plasticine "doughnut" at the bottom of the box to support the cranium upside-down. By remodeling the plasticine, the height of the cranium can be adjusted to match that of the millimeter paper, and a small spirit level on, for instance, the palate used to verify that it is parallel to the lens of the camera.

Every operator will have to adjust the photographic settings to their specific aim and study structure, planning everything in advance and anticipating all potential problems. This may be a tedious operation with much trial and error before the protocol is found. The careful settings of Cardini and Tongiorgi (2003) are one of the reasons why the TTD error in photographs of marmot hemi-mandibles was so small (Cardini, 2014), being about one-third and half that found in crania of plains zebras, marmots, and patas. In fact, photographs had a TTD error even smaller than in 2D shapes from high resolution scans of the same individuals (Cardini, 2014).

4.6 | What's next? The potential of 2D GMM for taxonomic assessment

A simple question inspired this study: Can 2D photographs of the ventral side of adult crania be used for a morphometric analysis of taxonomic variation in patas? The answer is "yes." The TTD error is similar, or even slightly smaller, than in previous analyses using the same methods on ventral crania of adults of some other species of mammal (Cardini, 2014; Cardini & Chiapelli, 2020). The magnitude of the TTD error, however, is not small when compared to differences among crania of adults of closely related species/subspecies. In these studies, it can be as large as one/fourth to one/fifth of individual variation, thus leading to a low percentage of 2D-3D sister replicates in the phenograms. Nonetheless, results of all taxonomic tests produced virtually identical results in 2D and 3D, which indicates that 2D GMM accurately captures patterns of group differences and should be appropriate for a quantitative assessment of morphological variability in patas.

We performed this study with two aims. The more general aim was to better understand whether and when 2D might be accurate despite the almost inevitable distortion of the third dimension in highly 3D structures such as mammal crania. The second, more specific, aim was to confirm (as we hoped) that we can now go on with a proper taxonomic study and, thus, assess the degree of variability in the adult crania across the geographic distribution of *Erythrocebus*. Most 2D studies consider the appropriateness of 2D GMM as a given. This, however, is often risky assumption. For example, what if 2D is inaccurate and one finds that out after an extensive data collection or, even later, after others fail to replicate 2D results using 3D landmarks?

Ventral crania provide a small piece of morphometric evidence. For an accurate taxonomic assessment we need to wait until results are corroborated by genetic data. Yet, for now, the 2D study we are preparing to carry out appears a promising approach to start improving our knowledge of the taxonomy of this genus of African primate. For robust morphometric findings on small taxonomic differences, however, we will need large samples representing all the main populations of patas. A recent GMM analysis of more than 4,000 adult crania from many genera of mammals (Cardini et al., 2021) suggests that a minimum number (within population and split by sex, for highly dimorphic taxa such as primates) might be in the range of 25–40 specimens. This number tends to, on average, produce estimates

WILEY 1429

of means and variance–covariance matrices close to those found in much larger samples of the same species and also strongly reduces errors in species identification based on cranial shape. As specimens are stored among many museums, photographing enough specimens is daunting, but not impossible. Visiting collections with patas crania will be expensive and time-consuming. Developing a standardized and reproducible protocol that yields minimal interoperator differences would allow others to obtain photographs from those museums that we are unable to visit. This collaboration would, thus, generate the largest samples.

Our study of patas has an interest that goes beyond the taxonomy and conservation of this group. Besides African primates (Butynski et al., 2013; Estrada et al., 2017), numerous mammals worldwide are in rapid decline and threatened with extinction (Ceballos et al., 2017; Schipper et al., 2008), some of which may hide more diversity than we presently recognize (Ceballos & Ehrlich, 2009). This partially cryptic variation, to some extent, reflects the choice of species definition and differences in taxonomic philosophy (e.g., Groves, 2001; Groves et al., 2017; Zachos, 2016, 2018). Yet, compared to the taxonomies presented in Wilson and Reeder (2005), the majority of the about 1,000 new species of mammal proposed in the recent taxonomic revision of Burgin et al. (2018) originates from the redefinition of previously known taxa based on updated scientific knowledge. Some of these new taxa may go unrecognized using different criteria and, therefore, lack robust support. For morphological and other phenotypic data there is also a chance that differences are plastic and do not reflect genetic divergence and adaptive change. Most likely, however, the majority of the "new" taxa represent unique components of biological diversity regardless of their taxonomic rank. Being the result of an irreproducible evolutionary history with generations of individuals selected by a variety of environmental pressures, preserving their genetic diversity could be key to enhancing their chance of survival in a rapidly changing, mostly deteriorating, natural landscape. With this variability, at least some populations might be resilient and able to adapt (Sgrò et al., 2011), and thus preserve diversity, but also phenotypic disparity and ecological functions (Mace, 2004; Mace & Purvis, 2008). Taxonomic uncertainties are just one of several problems in measuring biodiversity (Hortal et al., 2015) as more knowledge and conservation actions may not be enough to save species from extinction (Costello et al., 2013; Ellison, 2016; Kim & Byrne, 2006; Robinson, 2006), but without knowledge we will certainly end up losing diversity that we have never recognized (Kim & Byrne, 2006; Lees & Pimm, 2015; Tedesco et al., 2014).

AUTHOR CONTRIBUTIONS

Andrea Cardini: Conceptualization; data curation; study design, methodology and analysis; writing original draft, reviews and revisions. **Yvonne A. de Jong** and **Thomas M. Butynski**: Conceptualization; data curation; writing original draft, reviews and revisions.

ORCID

Andrea Cardini https://orcid.org/0000-0003-2910-632X
Yvonne A. de Jong https://orcid.org/0000-0002-8677-3738
Thomas M. Butynski https://orcid.org/0000-0002-0409-1515

REFERENCES

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2013). A field comes of age: Geometric morphometrics in the 21st century. *Hystrix, the Italian Journal of Mammalogy*, *24*, 7–14.

Alhajeri, B. H. (2018). Craniomandibular variation in the taxonomically problematic gerbil genus *Gerbillus* (Gerbillinae, Rodentia): Assessing the influence of climate, geography, phylogeny, and size. *Journal of Mammalian Evolution*, *25*, 261–276.

Álvarez, A., & Perez, S. I. (2013). Two- versus three-dimensional morphometric approaches in macroevolution: Insight from the mandible of caviomorph rodents. *Evolutionary Biology*, *40*, 150–157.

Armitage, K. B. (1999). Evolution of sociality in marmots. *Journal of Mammalogy*, *80*, 1–10.

Arnqvist, G., & Martensson, T. (1998). Measurement error in geometric morphometrics: Empirical strategies to assess and reduce its impact on measures of shape. *Acta Zoologica Academiae Scientiarum Hungaricae*, *44*, 73–96.

Bohórquez-Herrera, J., Aurioles-Gamboa, D., Hernández-Camacho, C., & Adams, D. C. (2017). Variability in the skull morphology of adult male California sea lions and Galapagos sea lions. In J. Alava (Ed.), *Tropical pinnipeds: Bio-ecology, threats and conservation* (pp. 22–49). CRC Press.

Bookstein, F. L. (2019). Pathologies of between-groups principal components analysis in geometric morphometrics. *Evolutionary Biology*, *46*, 271–302.

Borges, L. R., Maestri, R., Kubiak, B. B., Galiano, D., Fornel, R., & Freitas, T. R. O. (2017). The role of soil features in shaping the bite force and related skull and mandible morphology in the subterranean rodents of genus *Ctenomys* (Hystricognathi: Ctenomyidae). *Journal of Zoology*, *301*, 108–117.

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? *Journal of Mammalogy*, *99*, 1–14.

Buser, T. J., Sidlauskas, B. L., & Summers, A. P. (2018). 2D or not 2D? Testing the utility of 2D vs. 3D landmark data in geometric morphometrics of the sculpin subfamily Oligocottinae (Pisces; Cottoidea). *The Anatomical Record*, *301*, 806–818.

Butynski, T. M., & De Jong, Y. A. (2020). Taxonomy and biogeography of the gentle monkey Cercopithecus mitis Wolf, 1822 (Primates: Cercopithecidae) in Kenya and Tanzania, and designation of a new subspecies endemic to Tanzania. *Primate Conservation*, *34*, 71–127.

Butynski, T. M., Kingdon, J., & Kalina, J. (2013). *Mammals of Africa: Volume II. Primates*. Bloomsbury.

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

Cardini, A. (2014). Missing the third dimension in geometric morphometrics: How to assess if 2D images really are a good proxy for 3D structures? *Hystrix, the Italian Journal of Mammalogy*, *25*, 73–81.

Cardini, A. (2017). Left, right or both? Estimating and improving accuracy of one-side-only geometric morphometric analyses of cranial variation. *Journal of Zoological Systematics and Evolutionary Research*, *55*, 1–10.

Cardini, A. (2019). Integration and modularity in Procrustes shape data: Is there a risk of spurious results? *Evolutionary Biology*, *46*, 90–105.

Cardini, A. (2020a). Modern morphometrics and the study of population differences: Good data behind clever analyses and cool pictures? *The Anatomical Record*, *303*, 2747–2765.

Cardini, A. (2020b). Less tautology, more biology? A comment on "high-density" morphometrics. *Zoomorphology*, *139*, 513–529.

Cardini, A., & Chiapelli, M. (2020). How flat can a horse be? Exploring 2D approximations of 3D crania in equids. *Zoology*, *139*, 125746.

Cardini, A., & Elton, S. (2008a). Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. *Biological Journal of the Linnean Society*, *93*, 813–834.

Cardini, A., & Elton, S. (2008b). Variation in guenon skulls (I): Species divergence, ecological and genetic differences. *Journal of Human Evolution*, *54*, 615–637.

Cardini, A., & Elton, S. (2017). Is there a "Wainer's Rule"? Testing which sex varies most as an example analysis using GueSDat, the free Guenon Skull Database. *Hystrix, the Italian Journal of Mammalogy*, *28*, 147–156.

Cardini, A., Hoffmann, R. S., & Thorington, R. W. (2005). Morphological evolution in marmots (Rodentia, Sciuridae): Size and shape of the dorsal and lateral surfaces of the cranium. *Journal of Zoological Systematics and Evolutionary Research*, *43*, 258–268.

Cardini, A., Jansson, A., & Elton, S. (2007). A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. *Journal of Biogeography*, *34*, 1663–1678.

Cardini, A., & Loy, A. (2013). On growth and form in the computer era: From geometric to biological morphometrics. *Hystrix, the Italian Journal of Mammalogy*, *24*, 1–5.

Cardini, A., & Polly, P. D. (2020). Cross-validated between group PCA scatterplots: A solution to spurious group separation? *Evolutionary Biology*, *47*, 85–95.

Cardini, A., & Tongiorgi, P. (2003). Yellow-bellied marmots (*Marmota flaviventris*) "in the shape space" (Rodentia, Sciuridae): Sexual dimorphism, growth and allometry of the mandible. *Zoomorphology*, *122*, 11–23.

Carpenter, K. E., Sommer, H. J., & Marcus, L. F. (1996). Converting Truss interlandmark distances to Cartesian coordinates. In L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, & D. E. Slice (Eds.), *Advances in Morphometrics* (pp. 103–111). NATO ASI Series (A, Life Sciences).

Ceballos, G., & Ehrlich, P. R. (2009). Discoveries of new mammal species and their implications for conservation and ecosystem services. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 3841–3846.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E6089–E6096.

Chemisquy, M. A. (2015). Peramorphic males and extreme sexual dimorphism in *Monodelphis dimidiata* (Didelphidae). *Zoomorphology*, *134*, 587–599.

Chevret, P., Renaud, S., Helvaci, Z., Ulrich, R. G., Quéré, J.-P., & Michaux, J. R. (2020). Genetic structure, ecological versatility, and skull shape differentiation in *Arvicola* water voles (Rodentia, Cricetidae). *Journal of Zoological Systematics and Evolutionary Research*, *58*, 1323–1334.

Chiozzi, G., Bardelli, G., Ricci, M., De Marchi, G., & Cardini, A. (2014). Just another island dwarf? Phenotypic distinctiveness in the poorly known Soemmerring's gazelle, *Nanger soemmerringii* (Cetartiodactyla: Bovidae), of Dahlak Kebir Island. *Biological Journal of the Linnean Society*, *111*, 603–620.

Claude, J. (2008). *Morphometrics with R*. Springer Verlag.

Clauss, M., Nunn, C., Fritz, J., & Hummel, J. (2009). Evidence for a tradeoff between retention time and chewing efficiency in large mammalian herbivores. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *154*, 376–382.

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can we name Earth's species before they go extinct? *Science*, *339*, 413–416.

Cotterill, F. P. D., Taylor, P. J., Gippoliti, S., Bishop, J. M., & Groves, C. P. (2014). Why one century of phenetics is enough: Response to "are there really twice as many bovid species as we thought?". *Systematic Biology*, *63*, 819–832.

Cramer, J. S. (1987). Mean and variance of R^2 in small and moderate samples. *Journal of Econometrics*, *35*, 253–266.

Daboul, A., Ivanovska, T., Bülow, R., Biffar, R., & Cardini, A. (2018). Procrustes-based geometric morphometrics on MRI images: An example of inter-operator bias in 3D landmarks and its impact on big datasets. *PLoS One*, *13*, e0197675.

D'Anatro, A., & Lessa, E. P. (2006). Geometric morphometric analysis of geographic variation in the Río Negro tuco-tuco. *Ctenomys rionegrensis* (Rodentia: Ctenomyidae), *71*, 288–298.

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, *85*, 407–417.

De Jong, Y. A., & Butynski, T. M. (2017). Distributions in Uganda, Kenya, and North Tanzania of members of the Günther's dik-dik *Madoqua* (*guentheri*) and Kirk's dik-dik *M.* (*kirkii*) species groups, regions of sympatry, records of aberrant-coloured individuals, and comment on the validity of Hodson's dik-dik *M.* (*g.*) *hodsoni*. *Gnusletter*, *34*, 11–20.

De Jong, Y. A., & Butynski, T. M. (2020a). *Erythrocebus baumstarki. The IUCN Red List of Threatened Species 2020*: e. T92252436A92252442. https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T92252436A92252442.en

De Jong, Y. A. & Butynski, T. M. (2020b). *Erythrocebus patas* ssp. *pyrrhonotus. The IUCN Red List of Threatened Species 2020*: e. T92252480A92252486. https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T92252480A92252486.en

De Jong, Y. A., & Butynski, T. M. (2021). Is the southern patas monkey *Erythrocebus baumstarki* Africa's next primate extinction? Reassessing geographic distribution, abundance, and conservation. *American Journal of Primatology*, *83*(10). https://doi.org/10.1002/ajp.23316.

De Jong, Y. A., Rylands, A. B., & Butynski, T. M. (2020). *Erythrocebus patas. The IUCN Red List of Threatened Species 2020:* e.T174391079A17940998. https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T174391079A17940998.en

dos Reis, S. F., Duarte, L. C., Monteiro, L. R., & Von Zuben, F. J. (2002). Geographic variation in cranial morphology in *Thrichomys apereoides* (Rodentia: Echimyidae). II. Geographic units, morphological discontinuities, and sampling gaps. *Journal of Mammalogy*, *83*, 345–353.

Dryden, I. L. (2019). *Shapes package*. R Foundation for Statistical Computing.

Elliot, D. G. (1913). *A review of the primates. Vol. III: Anthropoidea (*Miopithecus *to* Pan*)*. Monograph Series. American Museum of Natural History.

Ellison, A. M. (2016). It's time to get real about conservation. *Nature*, *538*, 141.

Estrada, A., Garber, P. A., Rylands, A. B., Roos, C., Fernandez-Duque, E., Fiore, A. D., Nekaris, K. A.-I., Nijman, V., Heymann, E. W., Lambert, J. E., Rovero, F., Barelli, C., et al. (2017). Impending extinction crisis of the world's primates: Why primates matter. *Science Advances*, *3*, e1600946.

Evin, A., Baylac, M., Ruedi, M., Mucedda, M., & Pons, J. (2008). Taxonomy, skull diversity and evolution in a species complex of *Myotis* (Chiroptera: Vespertilionidae): A geometric morphometric appraisal. *Biological Journal of the Linnean Society*, *95*, 529–538.

Evin, A., Bonhomme, V., & Claude, J. (2020). Optimizing digitalization effort in morphometrics. *Biology Methods and Protocols*, *5*(1), bpaa023.

Evin, A., Souter, T., Hulme-Beaman, A., Ameen, C., Allen, R., Viacava, P., Larson, G., Cucchi, T., & Dobney, K. (2016). The use of close-range photogrammetry in zooarchaeology: Creating accurate 3D models of wolf crania to study dog domestication. *Journal of Archaeological Science: Reports*, *9*, 87–93.

Falkingham, P. L. (2012). Acquisition of high resolution three-dimensional models using free, open-source, photogrammetric software. *Palaeontologia Electronica*, *15*, 15pp.

Foote, M. (1997). The evolution of morphological diversity. *Annual Review of Ecology and Systematics*, *28*, 129–152.

Fornel, R., Cordeiro-Estrela, P., & De Freitas, T. R. (2010). Skull shape and size variation in *Ctenomys minutus* (Rodentia: Ctenomyidae) in geographical, chromosomal polymorphism, and environmental contexts. *Biological Journal of the Linnean Society*, *3*, 705–720.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Sage.

Fox, N. S., Veneracion, J. J., & Blois, J. L. (2020). Are geometric morphometric analyses replicable? Evaluating landmark measurement error and its impact on extant and fossil *Microtus* classification. *Ecology and Evolution*, *10*, 3260–3275.

Fruciano, C. (2016). Measurement error in geometric morphometrics. *Development Genes and Evolution*, *226*, 139–158.

Fruciano, C., Celik, M. A., Butler, K., Dooley, T., Weisbecker, V., & Phillips, M. J. (2017). Sharing is caring? Measurement error and the issues arising from combining 3D morphometric datasets. *Ecology and Evolution*, *7*, 7034–7046.

Giacomini, G., Scaravelli, D., Herrel, A., Veneziano, A., Russo, D., Brown, R. P., & Meloro, C. (2019). 3D photogrammetry of bat skulls: Perspectives for macro-evolutionary analyses. *Evolutionary Biology*, *46*, 249–259.

Gippoliti, S., & Rylands, A. B. 2020. *Erythrocebus poliophaeus. The IUCN Red List of Threatened Species 2020:* e. T164377509A164377626. https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T164377509A164377626.en

Groves, C. (2001). *Primate taxonomy*. Smithsonian Institution.

Groves, C. P., Cotterill, F. P., Gippoliti, S., Robovský, J., Roos, C., Taylor, P. J., & Zinner, D. (2017). Species definitions and conservation: A review and case studies from African mammals. *Conservation Genetics*, *18*, 1247–1256.

Grubb, P., Butynski, T. M., Oates, J. F., Bearder, S. K., Disotell, T. R., Groves, C. P., & Struhsaker, T. T. (2003). Assessment of the diversity of African primates. *International Journal of Primatology*, *24*, 1301–1357.

Hedrick, B. P., Antalek Schrag, P., Conith, A. J., Natanson, L. J., & Brennan, P. L. R. (2019). Variability and asymmetry in the shape of the spiny dogfish vagina revealed by 2D and 3D geometric morphometrics. *Journal of Zoology*, *308*, 16–27.

Hill, W. C. O. (1966). *Primates. Comparative anatomy and taxonomy. Volume 6: Catarrhini, Cercopithecoidea, Cercopithecinae*. Edinburgh University Press.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 523–549.

Ibrahim, K. M., Williams, P. C., Olson, A., Torounsky, R., Naser, E., Ghebremariam, F. H., & Masri, M. A. (2020). Genetic variation in morphologically divergent mainland and island populations of Soemmerring's gazelles (*Nanger soemmerringii*). *Mammal Research*, *65*, 403–412.

Isbell, L. A. (2013). *Erythrocebus patas* patas monkey (hussar monkey, nisnas). In T. M. Butynski, J. Kingdon, & J. Kalina (Eds.), *Mammals of Africa. Volume II: Primates* (pp. 257–264). Bloomsbury.

IUCN. (2020). *The IUCN Red List of Threatened Species 2020*. IUCN www.iucnredlist.org

Jojić, V., Blagojević, J., & Vujošević, M. (2011). B chromosomes and cranial variability in yellow-necked field mice (*Apodemus flavicollis*). *Journal of Mammalogy*, *92*, 396–406.

Katz, D., & Friess, M. (2014). Technical note: 3D from standard digital photography of human crania—A preliminary assessment. *American Journal of Physical Anthropology*, *154*, 152–158.

Kerhoulas, N. J., Gunderson, A. M., & Olson, L. E. (2015). Complex history of isolation and gene flow in hoary, Olympic, and endangered Vancouver Island marmots. *Journal of Mammalogy*, *96*, 810–826.

Kim, K. C., & Byrne, L. B. (2006). Biodiversity loss and the taxonomic bottleneck: Emerging biodiversity science. *Ecological Research*, *21*, 794–810.

Kingdon, J. (2013a). *Cercopithecus* (*nictitans*) group. Nictitans monkey group. In T. M. Butynski, J. Kingdon, & J. Kalina (Eds.), *Mammals of Africa. Volume II: Primates* (pp. 344–350). Bloomsbury.

Kingdon, J. (2013b). Genus *Madoqua* dik-diks. In J. Kingdon & M. Hoffmann (Eds.), *Mammals of Africa. Volume VI: Pigs, Hippopotamuses, Chevrotain, Giraffes, Deer and Bovids* (pp. 320–322). Bloomsbury.

Klenovšek, T., & Jojić, V. (2016). Modularity and cranial integration across ontogenetic stages in Martino's vole, *Dinaromys bogdanovi*. *Contributions to Zoology*, *85*, 275–289.

Klingenberg, C. P. (1998). Heterochrony and allometry: The analysis of evolutionary change in ontogeny. *Biological Reviews*, *73*, 79–123.

Klingenberg, C. P. (2008). Novelty and "homology-free" morphometrics: What's in a name? *Evolutionary Biology*, *35*, 186–190.

Klingenberg, C. P. (2011). MorphoJ: An integrated software package for geometric morphometrics. *Molecular Ecology Resources*, *11*, 353–357.

Klingenberg, C. P. (2013). Visualizations in geometric morphometrics: How to read and how to make graphs showing shape changes. *Hystrix, the Italian Journal of Mammalogy*, *24*, 15–24.

Klingenberg, C. P., Barluenga, M., & Meyer, A. (2002). Shape analysis of symmetric structures: Quantifying variation among individuals and asymmetry. *Evolution*, *56*, 1909–1920.

Kovarovic, K., Aiello, L. C., Cardini, A., & Lockwood, C. A. (2011). Discriminant function analyses in archaeology: Are classification rates too good to be true? *Journal of Archaeological Science*, *38*, 3006–3018.

Lalis, A., Evin, A., & Denys, C. (2009). Morphological identification of sibling species: The case of west African *Mastomys* (Rodentia: Muridae) in sympatry. *Comptes Rendus Biologies*, *332*, 480–488.

Lees, A. C., & Pimm, S. L. (2015). Species, extinct before we know them? *Current Biology*, *25*, R177–R180.

Loveless, A. M., Reding, D. M., Kapfer, P. M., & Papeş, M. (2016). Combining ecological niche modelling and morphology to assess the range-wide population genetic structure of bobcats (*Lynx rufus*). *Biological Journal of the Linnean Society*, *117*, 842–857.

Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *359*, 711–719.

Mace, G. M., & Purvis, A. (2008). Evolutionary biology and practical conservation: Bridging a widening gap. *Molecular Ecology*, *17*, 9–19.

Marcy, A. E., Hadly, E. A., Sherratt, E., Garland, K., & Weisbecker, V. (2016). Getting a head in hard soils: Convergent skull evolution and divergent allometric patterns explain shape variation in a highly diverse genus of pocket gophers (*Thomomys*). *BMC Evolutionary Biology*, *16*, 207.

Milenvić, M., Šipetić, V. J., Blagojević, J., Tatović, S., & Vujošević, M. (2010). Skull variation in Dinaric-Balkan and Carpathian gray wolf populations revealed by geometric morphometric approaches. *Journal of Mammalogy*, *91*, 376–386.

Mullin, S. K., & Taylor, P. J. (2002). The effects of parallax on geometric morphometric data. *Computers in Biology and Medicine*, *32*, 455–464.

Muñoz Muñoz, F., Quinto Sánchez, M., & González, J. R. (2016). Photogrammetry: A useful tool for three-dimensional morphometric analysis of small mammals. *Journal of Zoological Systematics and Evolutionary Research*, *54*, 318–325.

Murta-Fonseca, R. A., Machado, A., Lopes, R. T., & Fernandes, D. S. (2019). Sexual dimorphism in *Xenodon neuwiedii* skull revealed by geometric morphometrics (Serpentes; Dipsadidae). *Amphibia-Reptilia*, *40*, 461–474.

Myers, P., Lundrigan, B. L., Gillespie, B. W., & Zelditch, M. L. (1996). Phenotypic plasticity in skull and dental morphology in the prairie deer mouse (*Peromyscus maniculatus bairdii*). *Journal of Morphology*, *229*, 229–237.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, *82*, 591–605.

Napier, P. H. (1981). *Catalogue of primates in the British museum (natural history) and elsewhere in the British Isles. Part II: Family Cercopithecidae, subfamily Cercopithecinae*. British Museum (Natural History).

Navarro, N., & Maga, A. M. (2016). Does 3D phenotyping yield substantial insights in the genetics of the mouse mandible shape? *G3: Genes, Genomes, Genetics*, *6*, 1153–1163.

Nowak, K., Cardini, A., & Elton, S. (2008). Evolutionary acceleration and divergence in *Procolobus kirkii*. *International Journal of Primatology*, *29*, 1313–1339.

Oates, J. F., & Ting, N. (2015). Conservation consequences of unstable taxonomies: The case of the red colobus monkeys. In A. M. Behie & M. F. Oxenham (Eds.), *Taxonomic tapestries: The threads of evolutionary, behavioural and conservation research* (pp. 321–343). ANU Press.

Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Simpson, G., Solymos, P., Stevens, M., & Wagner, H. (2011). *Vegan: Community Ecology Package. R Package Ver. 2.0-2*.

Oxnard, C., & O'Higgins, P. (2009). Biology clearly needs morphometrics. Does morphometrics need biology? *Biological Theory*, *4*, 84–97.

Panchetti, F., Scalici, M., Carpaneto, G. M., & Gibertini, G. (2008). Shape and size variations in the cranium of elephant-shrews: A morphometric contribution to a phylogenetic debate. *Zoomorphology*, *127*, 69–82.

Pandolfi, L., Martino, R., Rook, L., & Piras, P. (2020). Investigating ecological and phylogenetic constraints in hippopotamidae skull shape. *Rivista Italiana di Paleontologia e Stratigrafia*, *126*, 37–49.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Robinson, J. G. (2006). Conservation biology and real-world conservation. *Conservation Biology*, *20*, 658–669.

Rohlf, F. J. (1998). On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology*, *47*, 147–158.

Rohlf, F. J. (2000). On the use of shape spaces to compare morphometric methods. *Hystrix, the Italian Journal of Mammalogy*, *11*, 1–17.

Rohlf, F. J. (2015). The tps series of software. *Hystrix, the Italian Journal of Mammalogy*, *26*, 9–12.

Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, *39*, 40–59.

Roth, V. (1993). On three-dimensional morphometrics, and on the identification of landmark points. In Marcus, Leslie, F., Elisa, B., & Antonio G.-V., (Eds.), *Contributions to Morphometrics. Museo Nacional de Ciencias Naturales, Madrid* (pp. 41–61). Editorial CSIC-CSIC Press.

Samuels, J. X. (2009). Cranial morphology and dietary habits of rodents. *Zoological Journal of the Linnean Society*, *156*, 864–888.

Scalici, M., Spani, F., Traversetti, L., Carpaneto, G. M., & Piras, P. (2018). Cranial shape parallelism in soft-furred sengis: Moving on a geographic gradient. *Journal of Mammalogy*, *99*, 1375–1386.

Schipper, J., Chanson, J. S., Chiozza, F., Cox, N. A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A. S., Stuart, S. N., Temple, H. J., Baillie, J., Boitani, L., Lacher, T. E., Jr., Mittermeier, R. A., Smith, A. T., Absolon, D., Aguiar, J. M., Amori, G., Bakkour, N., ... Young, B. E. (2008). The status of the world's land and marine mammals: Diversity, threat and knowledge. *Science*, *322*, 225–230.

Schlager, S. (2017). Morpho and Rvcg—Shape analysis in R. In G. Zheng, S. Li, & G. Szekely (Eds.), *Statistical shape and deformation analysis* (pp. 217–256). Academic Press.

Senn, H., Banfield, L., Wacher, T., Newby, J., Rabeil, T., Kaden, J., Kitchener, A. C., Abaigar, T., Silva, T. L., Maunder, M., & Ogden, R. (2014). Splitting or lumping? A conservation dilemma exemplified by the critically endangered dama gazelle (*Nanger dama*). *PLoS ONE*, *9*, e98693.

Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, *4*, 326–337.

Siberchicot, A., Julien-Laferrière, A., Dufour, A.-B., Thioulouse, J., & Dray, S. (2017). Adegraphics: An S4 lattice-based package for the representation of multivariate data. *The R Journal*, *9*, 198–212.

Solow, A. R. (1990). A randomization test for misclassification probability in discriminant analysis. *Ecology*, *71*, 2379–2382.

Souto, N. M., Murta Fonseca, R. A., Machado, A. S., Lopes, R. T., & Fernandes, D. S. (2019). Snakes as a model for measuring skull preparation errors in geometric morphometrics. *Journal of Zoology*, *309*, 12–21.

Tedesco, P. A., Bigorne, R., Bogan, A. E., Giam, X., Jézéquel, C., & Hugueny, B. (2014). Estimating how many undescribed species have gone extinct. *Conservation Biology*, *28*, 1360–1370.

Ting, N. (2008). Mitochondrial relationships and divergence dates of the African colobines: Evidence of Miocene origins for the living colobus monkeys. *Journal of Human Evolution*, *55*, 312–325.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Viscosi, V., & Cardini, A. (2011). Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. *PLoS One*, *6*, e25630.

Wallis, J. (2020). *Erythrocebus patas* ssp. *patas. The IUCN Red List of Threatened Species 2020*: e.T92252458A92252464. https://doi.org/10.2305/IUCN.UK.2020-2.RLTS.T92252458A92252464.en.

White, J. W., & Ruttenberg, B. I. (2007). Discriminant function analysis in marine ecology: Some oversights and their solutions. *Marine Ecology Progress Series*, *329*, 301–305.

Willmore, K. E., Zelditch, M. L., Young, N., Ah Seng, A., Lozanoff, S., & Hallgrímsson, B. (2006). Canalization and developmental stability in the Brachyrrhine mouse. *Journal of Anatomy*, *208*, 361–372.

Wilson, D. E., & Reeder, D. M. (2005). *Mammal species of the world: A taxonomic and geographic reference*. JHU Press.

Yazdi, F. T. (2017). Testing and quantification of cranial shape and size variation within *Meriones hurrianae* (Rodentia: Gerbillinae): A geometric morphometric approach. *Mammalian Biology*, *87*, 160–167.

Yazdi, F. T., Adriaens, D., & Darvish, J. (2012). Geographic pattern of cranial differentiation in the Asian midday jird *Meriones meridianus* (Rodentia: Muridae: Gerbillinae) and its taxonomic implications. *Journal of Zoological Systematics and Evolutionary Research*, *50*, 157–164.

Zachos, F. E. (2016). *Species concepts in biology: Historical development, theoretical foundations and practical relevance*. Springer.

Zachos, F. E. (2018). Mammals and meaningful taxonomic units: The debate about species concepts and conservation. *Mammal Review*, *48*, 153–159.

Zimmerman, D. W., Zumbo, B. D., & Williams, R. H. (2003). Bias in estimation and hypothesis testing of correlation. *Psicológica*, *24*, 133–158.

Zinner, D., Chuma, I. S., Knauf, S., & Roos, C. (2018). Inverted intergeneric introgression between critically endangered kipunjis and yellow baboons in two disjunct populations. *Biology Letters*, *14*, 20170729.

Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L. F., & Roos, C. (2013). Baboon phylogeny as inferred from complete mitochondrial genomes. *American Journal of Physical Anthropology*, *150*, 133–140.

APPENDIX A: THE MAIN ABBREVIATIONS SPECIFIC TO THIS STUDY (SEE MAIN TEXT FOR DETAILS)

2D: related to landmarks digitized on photographs. Each landmark is described by a pair of X and Y coordinates.

3D: related to landmarks digitized directly on crania. Each landmark is described by a triplet of X, Y and Z coordinates. Both for 2D and 3D, the coordinates can be raw (in mm, including differences in position and size) or Procrustes shape coordinates (after the superimposition has standardized size and positional differences).

  Configurations:

- FULL-0: all 25 landmarks digitized twice on the photographs in order to assess 2D digitization error; results from this dataset are emphasized using a light gray background in the tables.
- FULL-1: all 25 landmarks digitized on the photographs (averages of the two replicates of FULL-0) and re-digitized directly on the 3D crania.
- RED-2: reduced configuration using the same data as in FULL-1 but after removing landmarks 11, 15, 16, and 18 (and the corresponding mirror-reflected paired landmarks).

- RED-3: same as FULL-1 after removing landmarks 7, 11, and 23 (and the corresponding mirror-reflected paired landmarks).
- RED-4: same as FULL-1 after removing landmarks 11, 15, 16, 18, and 23 (and the corresponding mirror-reflected paired landmarks).

GMM: geometric morphometrics. Here, specifically referred to the set of methods based on the Procrustes superimposition of anatomical landmarks.

ME: measurement error.

Variance metrics used for multivariate shape data:

- VAR1: the sum of variances of the shape coordinates.
- VAR2: the mean of pairwise Procrustes shape distances among all individuals in a sample.
- VAR3: the 90th percentile of the same set of pairwise Procrustes distances used in VAR2.

varcov: matrix of variances and covariances of the Procrustes shape coordinates.

TTD: 2D to 3D approximation.

XbgPCA: cross-validated between group PCA (principal component analysis).