A geometric morphometric approach to the study of variation of shovel-shaped incisors.

Journal:	*American Journal of Physical Anthropology*
Manuscript ID	AJPA-2018-00075.R2
Wiley - Manuscript type:	Technical Note
Date Submitted by the Author:	n/a
Complete List of Authors:	Carayon, Delphine; Universite Toulouse III Paul Sabatier Faculte des Sciences et d'Ingenierie, Laboratoire AMIS, UMR 5288 CNRS; Universite de Montpellier Faculte d'Odontologie, Service de Prothèses Adhikari, Kaustubh; University College London, Department of Genetics, Evolution and Environment, and UCL Genetics Institute Montsarrat, Paul; Universite Toulouse III Paul Sabatier Faculte des Sciences et d'Ingenierie, Faculté d'Odontologie Dumoncel, Jean; Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, UMR 5288 CNRS-Université de Toulouse (Paul Sabatier), Braga, José; AMIS UMR 5288 CNRS, Hominid Evolutionary Biology; University of the Witwatersrand, Institute for Human Evolution Duployer, Benjamin; Centre Inter-universitaire de Recherche et d'Ingénierie des Matériaux, UMR 5085 CNRS-Université de Toulouse (Paul Sabatier) Delgado, Miguel; Consejo Nacional de Investigaciones Cientificas y Tecnicas Fuentes-Guajado, Macarena; University College London Research Department of Genetics Evolution and Environment, Department of Genetics, Evolution and Environment de Beer, Frikkie; South African Nuclear Energy Corporation Hoffman, Jakobus; Pelindaba, Radiation Science Oettlé, Anna; University of Pretoria, Department of Anatomy Donat, Richard; Institut National de Recherches Archéologiques Préventives, St Estève Pan, Lei; Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, ; Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, UMR 5288 CNRS-Université de Toulouse (Paul Sabatier), Ruiz-Linares, Andres; University College London, Biology Tenailleau, Christophe; Centre Inter-universitaire de Recherche et d'Ingénierie des Matériaux, UMR 5085 CNRS-Université de Toulouse (Paul Sabatier) Vaysse, Frédéric; Universite Toulouse III Paul Sabatier Faculte des Sciences et d'Ingenierie, Faculté d'Odontolgie Esclassan, Rémi; Universite Toulouse III Paul Sabatier Faculte des Sciences et d'Ingenierie, Faculté d'Odontologie Zanolli, Clément; Universite Toulouse III Paul Sabatier, Laboratoire AMIS,

	UMR 5288
Key Words:	shovel-shape incisors, ASUDAS, Procrustes and non-Procrustes superimpositions, virtual anthropology
Subfield: Please select 2 subfields. Select the main subject first.:	Human biology [living humans; behavior, ecology, physiology, anatomy], Primate biology [behavior, ecology, physiology, anatomy]

A geometric morphometric approach to the study of variation of shovel-shaped incisors

Delphine Carayon[1,2], Kaustubh Adhikari[3], Paul Monsarrat[4,5], Jean Dumoncel[1], José Braga[1,6], Benjamin Duployer[7], Miguel Delgado[8,9], Macarena Fuentes-Guajardo[3], Frikkie de Beer[10], Jakobus W. Hoffman[10], Anna C. Oettlé[11,12], Richard Donat[1,13], Lei Pan[14,15], Andres Ruiz-Linares[3,16,17], Christophe Tenailleau[7], Frédéric Vaysse[1,4], Rémi Esclassan[1,4], Clément Zanolli[1].

[1]UMR 5288 CNRS, Université Toulouse III - Paul Sabatier, France

[2]Faculté d'Odontologie, Université Montpellier I, France

[3]Department of Genetics, Evolution and Environment, and UCL Genetics Institute, University College London, London, United Kingdom

[4]Faculté d'Odontologie, Université Toulouse III - Paul Sabatier, France

[5]STROMALab, CNRS ERL 5311, EFS, INP-ENVT, Inserm, Université Toulouse III - Paul Sabatier, France

[6]Evolutionary Studies Institute and School of Geosciences, University of the Witwatersrand, PO WITS, Johannesburg 2050, South Africa

[7]Centre Inter-universitaire de Recherche et d'Ingénierie des Matériaux, UMR 5085 CNRS, Université de Toulouse III - Paul Sabatier, France

[8]Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Buenos Aires, República Argentina

[9]División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, República Argentina

[10]Radiation Science Department, South African Nuclear Energy Corporation (Necsa), Pelindaba, South Africa

[11]Department of Anatomy, University of Pretoria, South Africa

[12]Department of Anatomy, Sefako Makgatho Health Sciences University, South Africa

[13]Institut National de Recherches Archéologiques Préventives- St Estève, France

[14]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, CAS, Beijing, China

[15]State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, CAS, Nanjing, China

[16]Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai, China

[17]Laboratory of Biocultural Anthropology, Law, Ethics, and Health (Centre National de la Recherche Scientifique and Etablissement Français du Sang, UMR-7268), Aix-Marseille University, Marseille, France

Corresponding author: Delphine Carayon (delphinedom.carayon@orange.fr)

Keywords: shovel-shape incisors; ASUDAS; Procrustes and non-Procrustes superimpositions; virtual anthropology

Abstract

Objectives: The scoring and analysis of dental non-metric traits are predominantly accomplished by using the Arizona State University Dental Anthropology System (ASUDAS), a standard protocol based on strict definitions and three-dimensional dental plaques. However, visual scoring, even when controlled by strict definitions of features, visual reference and the experience of the observer, includes an unavoidable part of subjectivity. In this methodological contribution, we propose a new quantitative geometric morphometric approach to quickly and efficiently assess the variation of shoveling in modern human maxillary central incisors (UI1).

Materials and Methods: We analyzed 87 modern human UI1s by means of virtual imaging and the ASU-UI1 dental plaque grades using geometric morphometrics by placing semilandmarks on the labial crown aspect. The modern human sample was composed of individuals from Europe, Africa and Asia and included representatives of all seven grades defined by the ASUDAS method.

Results: Our results highlighted some limitations in the use of the current UI1 ASUDAS plaque, indicating that it did not necessarily represent an objective gradient of expression of a non-metric tooth feature. Rating of shoveling tended to be more prone to intra- and inter-observer bias for the highest grades. In addition, our analyses suggest that the observers were strongly influenced by the depth of the lingual crown aspect when assessing the shoveling.

Discussion: In this context, our results provide a reliable and reproducible framework reinforced by statistical results supporting the fact that open scale numerical measurements can complement the ASUDAS method.

Introduction

Teeth display morphological variations of the crown and roots that differ substantially among modern human and fossil groups, some dental characteristics being predominant in certain groups or populations (Turner et al., 1991). As stated by Hillson, "human eyes and brain are unsurpassed in discerning tiny differences between objects compared side by side, but it is much more difficult to define a scheme for recording size and shape in such a way that comparisons can be made between hundreds of such objects" (Hillson, 1996: 68). For this reason, since the 19th century, several attempts have been made to classify and assess differences between fossil and extant human populations, at first using detailed descriptive approaches and later elaborating scoring systems (reviewed in Irish and Scott, 2016).

Following the influential early works of Hrdlička (1920) and Dahlberg (1956), who standardized a four grade classification plaque for upper incisor shoveling, some researchers tried to reduce the visual subjectivity by measuring the depth of the lingual fossa. However, they had little success because of issues with the precision of the method (Dahlberg and Mikkelsen, 1947, Carbonell, 1963; Goaz and Miller, 1966; Hanihara, 1969). Later Scott (1973) developed an eight degree scale that was then adapted and integrated by Turner and collaborators (1991) into a formal system for scoring non-metric aspects of dental morphology: the Arizona State University Dental Anthropology System (ASUDAS) (Scott, 1973; Turner et al., 1991; Scott and Turner, 1997). This widely-used standard protocol is based on reference plaster plaques representing the casts of selected teeth showing a gradient of expression of a particular trait (Turner et al., 1991; Scott and Turner, 1997; Scott et al., 2018). Since their initial publication, the number of traits and plaques have increased and some of them have been adapted to the range of variation expressed by fossil hominins (Bailey, 2006; Bailey and Hublin, 2013; Irish et al., 2013; Irish and Scott, 2016).The scoring and analysis of dental non-metric traits currently represents a common diagnostic procedure to highlight ancestry/genetic affinities and investigate human variation in anthropological, paleoanthropological and forensic studies (Turner et al., 1991; Scott and Turner, 1997; Irish, 1998; Irish, 2014a; Irish and Guatelli-Steinberg, 2003; Bailey and Hublin, 2013; Zanolli, 2013; Zanolli et al., 2014; Irish and Scott, 2016). If the observer has been trained by an expert, the ASUDAS approach to morphological characters is easy, fast and reliable, and remains the gold standard technique today (Scott and Irish, 2017; Scott et al., 2018). However, visual scoring, even when controlled by strict definitions of features and the experience of the observer, includes an unavoidable part of subjectivity. In fact, the

assessment of shoveling defined by the ASUDAS method has some major limitations inherent to its definition. The specimens selected to develop the reference grades on the plaque were chosen by qualitative appreciation, which does not necessarily represent the morphological variation in a linear way. This can lead to minimal visual difference between some grades of expression and so to the difficulty experienced by users in classifying the analyzed specimens with regard to the ASUDAS (especially for beginners). In brief, both the selection of the reference teeth when creating the ASUDAS method and the comparison of the dental specimens with the ASUDAS plaques are dependent on observations/palpations and the experience of the observer (i.e., dependent on operator subjectivity). Nichol and Turner (1986) have shown that the intra-observer error when assessing the expression of incisor shoveling is small: 4.1% for more than 1 grade difference and only 2% for presence/absence differences. However, as mentioned by Scott and Turner (1997), "it will probably never be possible to attain 100% concordance in replicated observations of tooth crown and root traits, either by single observers or between observers. The reference plaques developed by Dahlberg, K. Hanihara, Turner, and others have enhanced observational precision but they have not been a panacea for the reasons noted above (i.e., threshold expressions, post-eruptive modifications, surficial noise, varying levels of experience, etc.)" (Scott and Turner, 1997:72).

Incisor shoveling is one of the non-metric features that has received the most attention from anthropologists as an indicator of relationships among populations and it is frequently used for its taxonomic and phylogenetic relevance (e.g., Scott and Turner, 1997; Bailey and Hublin, 2013; Irish et al., 2013, 2014; Carter et al., 2014). This feature can be defined as the degree of elevation of the mesial and distal lingual marginal ridges on the lingual surface of the maxillary incisors, canines and mandibular incisors, with more pronounced forms enclosing a fossa (Hrdlička, 1920; Dahlberg, 1956; Turner et al., 1991, Scott and Turner, 1997). Shoveling is more marked and variable in the upper central and lateral incisors, the former being the polar tooth (Irish and Scott, 2016). Ales Hrdlička (1920) was the first to classify the degree of expression of shovel shaped incisors, assess this variation among several human populations and describe its occurrence in non-human species (Scott and Turner, 1997). Among his findings, he indicated that the prevalence and expression of incisor shoveling showed marked geographic variation in modern human populations, being frequent and strongly expressed in Asia, with a South to North increasing cline, but less frequent and weaker in Africa and Europe (Mizogushi, 1985; Turner, 1990; Kimura et al., 2009). Some workers have attempted to quantify the degree of development of the shoveling with an interval scale. Dahlberg and Mikkelsen (1947) used a Vernier scale with a modified Boley

Gauge to measure the depth of the incisor lingual fossa in millimeters. Hanihara et al. (1975) measured lingual fossa depth in a Japanese population in order to obtain metrical data to calculate intrafamilial correlations. Taverne et al. (1979) tried to measure various parts of a tooth crown surface by an indirect three-dimensional measurement method using photogrammetry and a Moiré pattern (Mizoguchi, 1985). Also, in a shovel-shaped tooth, the marginal ridges may extend from the incisal edge to the basal eminence and sometimes, in very pronounced cases, the ridges can converge on the eminence. In addition, the two marginal ridges may exhibit different degrees of expression (Mizoguchi, 1985). However, according to Scott and Turner (1997), the mesial and distal marginal ridges are so strongly correlated that they can be considered together as a single trait (Scott and Irish, 2017). Crummett also tried to summarize the main characteristics of incisor shoveling by considering three aspects: the expression of the marginal ridges, the development of a lingual tubercle at the lingual base of the crown, from a small swelling to an independent cusp, and the crown convexity or curvature (Crummett, 1994, 1995). More recently, using X-ray microtomographic imaging, Denton investigated the relationship of these three aspects between the external surface of the incisor crown and the enamel-dentin junction in a limited sample of 10 extant humans (Denton, 2011).

Although the expression of dental non-metric features may be sensitive to environmental or epigenetic factors (Mizoguchi, 2013), it is predominantly determined by genetic factors (Scott and Turner, 1997; Jernvall et al., 2000; Salazar-Ciudad and Jernvall, 2002, 2010; Park et al., 2012). To date, the best known genetically-correlated dental trait is incisor shoveling, which involves a single nucleotide polymorphism (SNP) of the ectodysplasin A receptor gene (*EDAR*), the most likely target of positive selection in Asian populations resulting in marked shovel shaped teeth (Kimura et al., 2009, 2012). However, *EDAR* has pleiotropic effects and a recent study suggested that it was selected in Asian groups for its effect of increasing ductal branching in the mammary gland, thereby amplifying the transfer of critical nutrients to infants via the mother's milk (Hlusko et al., 2018). In this case, the dental phenotypic expression associated with this gene could simply represent a side effect. In any case, shoveling constitutes a critical marker to discriminate between human groups and assess ancestry.

The objective of this contribution is to propose a new and complementary quantitative methodological approach to study the concavity of the palatal surface of UI1, used here as a proxy for the variation of the degree of expression of shoveling. We elaborate a geometric morphometric (GM) method taking the depth and shape of the labial incisor crown aspect

(i.e., two of the three aspects of shoveling: the expression of the marginal ridges and the curvature of the lingual aspect) into account to assess the degree of UI1 shoveling on a continuous scale. After comparison with the classical ASUDAS method, we discuss the implications of implementing such geometric morphometric analyses for the study of the modern human variability of dental traits and to better track evolutionary trends in hominins.

Material and methods

Sample and scanning procedures

Our sample consisted of 87 modern human permanent maxillary central incisors (UI1). It included specimens of European (n=44), South African (n=30) and Chinese (n=13) ancestry, as listed in Table 1. Only unworn to moderately worn tooth crowns (reaching maximum stage 2 as defined by Smith, 1984, and corresponding to a thin line of dentine exposure) having no particular damage or pathology on the labial aspect were included in the analyses. Visual scoring of the 87 specimens was achieved by two observers (DC and CZ) following the ASUDAS method (Supporting Information Table 1).

We also analyzed the original ASUDAS UI1 shoveling (ASU-UI1) plaque based on Dahlberg's work (1956) and developed by Turner and collaborators (1991). This plaque includes seven grades of shoveling expression, from the weakest (grade 0) to the most marked (grade 6) (Supporting Information Table 2). In a recent revision of the ASUDAS method, Scott and Irish (2017) described an 8th stage for UI1 shoveling (grade 7, defined as any expression that exceeds grade 6, involving marginal ridges that fold around on themselves, similar to grade 6 on the UI2 shoveling plaque) but they did not find any suitable example to put on the plaque. For this reason, we did not consider this last grade here.

The 44 European specimens were scanned by X-ray microtomography (micro-CT) at the CIRIMAT facility of the University of Toulouse with a Phoenix/GE Nanotom 180 instrument, using the following parameters: 100 kV, 100 µA, 0.36° angular step. The virtual records were reconstructed to a voxel size of 22 to 25 µm. The 30 South African teeth and the reference plaque ASU-UI1 were scanned by X-Ray Micro-CT at the MIXRAD facility of the South African Nuclear Energy Corporation SOC Limited (Necsa), with a Nikon XT H225-L instrument by using similar parameters, and reconstructed to a voxel size ranging from 42 to 50 µm (Hoffman and de Beer, 2012). The 13 modern human Asian teeth were scanned by X-Ray micro-CT using similar parameters at the Institute of Vertebrate Paleontology and Paleoanthropology of Beijing, China, and reconstructed to a voxel size of 31.4 µm.

Data were imported into the 3D analytical software Avizo v.8.0. (FEI Visualization Sciences Group) so that 3D renderings of the tooth external surface could be visualized and processed. Teeth were first segmented semi-automatically by using a thresholding approach (Spoor et al., 1993; Fajardo et al., 2002; Coleman and Colbert, 2007) and a surface was generated from the segmented object. The maximum of curvature was measured on each UI1 crown surface using the "MaxCurvature" module of Avizo. This allowed us to determine the extreme curvature line of the mesial and distal lingual crests and use these maxima as starting and ending points of our GM analyses. The cervical best fit plane was defined by placing at least three landmarks at the most apical points of the cervix on the labial and palatal aspects (points of maximum curvature on the labial and lingual sides of the cement enamel junction; Le Cabec et al., 2013). We translated this reference plane to the midpoint between the most incisal and the most cervical points of the crown (Figure 1a) and then placed 100 semilandmarks along this middle plane following the curve of the lingual aspect of the crown (Figure 1b).

Statistical analysis

Intra-observer error (reliability) of the ASUDAS visual scoring was assessed with respect to the UI1 reference plaque by intra-class correlation (ICC) using a two-way mixed effects “absolute agreement” model (Koo and Li, 2017). ICC is generally used to assess the correlation of various units organized in groups and describes how strongly units in the same group resemble each other. This analysis was done in order to check for both consistency (also referred to as precision in the literature; e.g., if a tooth is actually ASUDAS category 2, but two raters independently assign it to category 5, they are highly consistent with each other but they have a large bias of 3 units; Schrout and Fleiss, 1979; Joint Committee for Guides in Metrology, 2008; Hughes and Hase, 2010) and accuracy (i.e., looking for the degree of bias/error between observers and our objective landmark-based method, e.g., if grade 3 actually corresponds to grade 3 plaque).

To objectively compare the degree of concavity of the labial surface of each incisor from our sample with the grades of the reference plaque, we performed Procrustes analyses of the semilandmarks. In the Procrustes method, the original landmarks from all samples are first superimposed and aligned with one another to produce the Procrustes coordinates. Subsequently, a principal component analysis (PCA) of the Procrustes coordinates is performed.

The reliability of this computer-based technique was assessed by intra-class correlation (ICC) of the 100 landmark coordinates among the three operators and 30 samples. Reliability was higher when the distance between the landmarks assigned by two raters on the same sample was small. We considered the distance between the landmark and the origin of the 3D orthonormal reference as outcome, together with the individual X, Y and Z float coordinates. ICC was obtained after a two-way random effects "absolute agreement" model (Koo and Li, 2017). Levels of agreement between raters were also visually appreciated using Bland-Altman plots. (This kind of plot, assessing the degree of agreement between two observers, is similar to a Tukey mean-difference plot.)

We also superimposed the curves in a non-Procrustes way, aligning the first and last point (0 and 100 respectively) of each curve (Figure 2a). This alignment procedure requires an initial 3D rotation step, which is similar to Procrustes methods, but the subsequent steps are different from Procrustes. Since each curve lies in an approximate 3D plane (the semilandmarks are placed along a plane, see Figure 1b), the curve is rotated to approximately align with the X-Y plane (equation of the 3D plane is obtained by fitting a linear regression on the X, Y & Z coordinates, as the equation of a 3D plane is $aX + bY + cZ + d = 0$). After this alignment, the Z coordinate is discarded. The 2D data is then rotated again so that the first and last points lie on the Y axis (i.e. X coordinate = 0). Finally, they are scaled by a constant factor on both axes to achieve a fixed Y-axis range of 1, i.e. the first and last points of each curve now have coordinates (0,0) and (0,1) respectively. This scaling step is a major difference from the Procrustes method, as the classical Procrustes method scales to have a centroid size of 1, but here the curves are scaled to have a Y-axis range of 1 to facilitate comparisons. We then used these aligned coordinates to measure two metrics, the maximum depth of the lingual aspect with respect to the first and last points of the curves and the hollow area of the curves (Figure 2b). These metrics are not data-dependent like the previous one (PC scores have to be calculated in the whole sample, and values change according to the sample composition) as both the depth and hollow area can be measured directly on the aligned landmarks. Principal components were also calculated from the non-Procrustes aligned coordinates (X & Y).

All statistical analyses and graphic data visualization were performed in MATLAB R2017b (MATLAB and Statistics Toolbox Release, R2017b) and R 3.4.3 (R Core Team, 2018). The following R packages were used: scatterplot3d (Ligges and Mächler, 2003), shapes (Dryden, 2017), ade4 (Dray and Dufour, 2007), irr (Gamer et al., 2012) and Bland Altman Leh (Lehnert, 2015).

Most morphological variations in the human dentition vary on a continuous scale. However, for simplicity of representation, dental anthropological assessment schemes often use two or more categories into which the range of variation is 'binned' or 'categorized'. For instance, the amount of melanin pigmentation in the eye is a continuous quantity, but in traditional analyses it has been categorized into blue vs. brown to represent absence/presence of melanin, and historically considered to be a Mendelian trait until modern quantitative analysis showed its complex polygenic nature. Scott and Turner (1997) noted, specifically with the example of incisor shoveling, that such nonmetric dental traits are possibly 'quasi continuous' (ordinal or dichotomous) traits, derived from an underlying continuous trait. For example, while the depth of the incisor crown is a continuous quantity, it can be dichotomized into absence/presence indicating whether the amount of curvature is below a certain threshold. In such cases, the underlying continuous variable is called a 'latent variable' corresponding to the assessed categorical variable. Our analyses suggested that the maximum depth metric was the most likely candidate for any underlying 'latent' quantitative variable that might be the basis of the ASUDAS categories for shoveling (see Results below). The results also suggested that the relationship between the maximum depth variable and the ASUDAS categories was monotonic but non-linear, i.e. when the latent variable increases the categories also increase, but the spacing between categories is unequal. Therefore, to 'predict' an objective ASUDAS category for the 87 modern human specimens, we constructed a prediction function using the maximum depth values and numerical categories of the ASUDAS specimen teeth. To preserve nonlinearity, a spline function was fitted on these values, which was then used as the interpolant to obtain predicted ASUDAS categories from the 3-D measured maximum depth values on the 87 modern human specimen casts. The predicted ASUDAS categories were allowed to contain decimals to retain more precision, instead of rounding them off to the nearest integer category (e.g., 1.67 instead of 2). Similarly, some of the observer-assigned categories that had an intermediate rating (0_1 meaning a category between ASUDAS references 0 and 1) were allowed to retain them (the rating 0_1 was assigned the middle value of 0.5, for example). These objective predicted values were compared with subjective observer-assigned rating values via ICC to obtain accuracy (i.e. unbiasedness) measurements.

Results

Our ICC intra- and inter-observer tests on the visual scoring showed highly consistent assessment of shoveling within and between raters (Table 2). Consistency was high for the whole sample, but also for each chrono-geographic sub-sample considered in this study. In

contrast, the accuracy of the visual assessment performed by the raters, measured using the ICC between the rater-assigned category and the predicted category was moderate for the Europeans and Africans. For the Chinese sample, the accuracy of the visual assessment was very low (further discussion below).

We then looked at the results of the landmark-based analyses. Following the standard procedure in geometric morphometrics, principal component analysis (PCA) was used to explore the morphospace. Principal components (PCs) were calculated using the Procrustes coordinates of all landmarks, for all samples including the ASUDAS reference casts. Plotting the top PCs enabled us to visualize the morphospace, to see how the samples were distributed with respect to the ASUDAS grades (Figure 3). The first two components (PC1 and PC2) of the PCA accounted for 87.4% of the total variance (81.5% for PC1 and 5.9% for PC2).

PC shape changes could be visualized by plotting the PC loadings. PC loadings for the Procrustes and non-Procrustes methods were very similar; non-Procrustes PC shape changes are shown in Supporting Information Figure 1. While PCs are more difficult to interpret than direct measurements, the PC shape changes give us some idea of the morphological aspects they capture. The shape of PC1 is roughly proportional to the curve of the lingual aspect, i.e. the lingual fossa has the highest weight. Therefore the deeper the lingual fossa is from the baseline, the higher the PC value is. This explains why the PC has such high correlations with the maximum depth metric (Table 3). And since the deeper the lingual fossa is, the more the marginal ridges protrude with respect to the fossa, PC1 is also proportional to the ASUDAS shoveling grade (Table 4).

The shape of PC2 gives greatest weight to the corners of marginal ridges, thereby being proportional to the angle between the labial palate and the marginal ridges. This angle is called ‘labial convexity’ in Denton (2011), where it is shown that its relationship with the ASUDAS shoveling grade is not monotonic (largest angles for grades 2-3), which explains why PC2 is not strongly correlated with it (Table 4).

The shape of PC3 seems to reflect the left-right asymmetry present in the shape of the curve, in particular the asymmetry in the two angles. As asymmetry of the ridges is not relevant in the definition of the ASUDAS grades, this explains why PC3 is not correlated with the grades either (Table 4).

The shapes of later PCs, such as PC4, are much harder to interpret and, given that they capture a very small fraction of the variability, they might simply reflect random statistical variation.

Along PC1, the European (French contemporary and medieval) and South African specimens showed a similarly reduced expression of shoveling (expressed here by reduced lingual depth and a more linear morphology), overlapping with the ASUDAS grades 0 and 1, while the Chinese material encompassed the grades 2 to 5. Even though the ASUDAS grades tend to follow a trend along PC1, their distribution is not linearly organized and is heterogeneous (Figure 3a and Figure 4a). Grades 0-1, 2-3, 4-5 and 6 tended to form four clusters and grades were not equidistant from one another. There was no visible discrimination between the chrono-geographic human samples and the ASUDAS grades along PC3 and PC4, which represented 3.16% and 2.55%, respectively, of the total variance (Figure 3b). We tested the reproducibility of this Procrustes method. Our results show that the positioning of the landmarks was highly reproducible, with an ICC >0.990. The graphical Bland-Altman method confirmed this high level of agreement (Figure 5).

When considering the Non-Procrustes analysis, similar results were obtained. In Figure 4, the histograms showing the distribution of maximum depth (Figure 4b) and hollow area (Figure 4c) of the specimens and the ASUDAS grades also highlight the non-linear scattering of the reference grades. Again, grades 0 and 1 are close to each other, while grades 2 to 5 are grouped together and 6 is alone. In accordance with the knowledge that East Asian populations show a higher degree of shoveling than the rest, the histogram of Chinese samples for maximum depth (Figure 4b) shows no overlap with the French and South African samples. It is also interesting to note that several samples from each population have a maximum depth value that is intermediate between the ASUDAS categories 1 & 2 (which have a large gap between them). This might create some difficulties for the observers to assign ASUDAS ratings to them, and also reduce the distinction between populations when compared via ASUDAS category frequencies. However, the quantitative maximum depth measurements provide a complete separation of the Chinese samples from the rest. The hollow area metric also achieves near-complete separation and shows a similar pattern.

Principal components calculated from the non-Procrustes aligned coordinates showed trends similar to those of the Procrustes PCs. PC1 alone explained 92% of the variation, while PC2 explained a further 2.7%. The PC loadings are represented as PC shape changes in Supporting Information Figure 1.

Table 3 presents the correlation between these various metrics calculated on the 87 human specimens plus the 7 ASUDAS reference casts. In addition to the two direct measurements, maximum depth and hollow area, PC1 and PC2 are used from both Procrustes and non-Procrustes aligned methods. This indicates that the two direct metrics are very similar to each

other, so it is sufficient to use either one. It also shows that both direct metric measurements, maximum depth and hollow area, have very high correlation with PC values obtained from the two analyses, as expected from the observations above, and therefore can be used in lieu of Procrustes PC values. The advantage of using direct measurement metrics over Procrustes PCs is that they are more directly interpretable, and not dependent on the whole dataset.

The correlation of ASUDAS category values with these metrics calculated on the ASUDAS reference casts was also assessed (Supporting Information Table 3). It shows that both metrics, as well as PC1 from both analyses, are highly correlated with the category values, while maximum depth has the highest correlation (98%). It also indicates that the Procrustes and non-Procrustes PCs are very similar to each other. The major feature of the ASUDAS scale for shoveling is the progressively increasing expression of the marginal ridges (Scott and Turner, 1997; Scott and Irish 2017). Conversely, more protruding ridges imply a deeper lingual fossa, and hence the maximal depth metric gives a measure of the development of the mesial and distal ridges and constitutes an appropriate metric to quantitatively evaluate the expression of shoveling.

When combining the correlation patterns with the observation above that the maximum depth metric provides a complete separation of the Chinese samples from the rest, we considered the maximum depth to be the most likely candidate for any underlying 'latent' quantitative variable that might be the basis of the ASUDAS categories. Considering that maximum depth has a 99% correlation with non-Procrustes PC1, which explains 92% variation of the landmark coordinates, it can be said that the maximum depth metric captures a large part of the morphological variation of the labial aspect of the UI1 crown.

To assess the accuracy of the visual ASUDAS assessments made by two observers, we needed to compare them with objective 'true' assessments of the modern human specimens. To make such objective estimates of ASUDAS categories on the specimens, we started with the objectively measured maximum depth metric, and employed the previously described prediction function. The prediction procedure uses the spline interpolation function constructed entirely on the basis of the ASUDAS reference plaque (Figure 6), and is therefore free of subjectivity arising from other human raters. The spline model compares the maximum depth measurements against ASUDAS categories for the ASUDAS reference plaque elements. Using this objective prediction function on the objectively measured maximum depth metric for a new specimen provides an objective estimate of the ASUDAS grade for the new specimen. This estimated grade can then be compared to the subjective ASUDAS ratings provided by a human observer in order to assess the accuracy of the observer.

As found in our ICC intra- and inter-observer tests (Table 2), the specimens the observers assessed as expressing low degrees of shoveling (grades 0-1), namely most of the South African and French samples, were closer to the predicted 'true' values (using the ASUDAS reference grades), being less scattered around the spline interpolation curve used for the prediction (Figure 6). In contrast, observer ratings for those recorded as having marked shoveling (grade 2 and above) – primarily the Chinese samples – were much more scattered around the spline curve, indicating that they differed more from the predicted 'true' ASUDAS grades (Figure 6). This corroborates our observation above that the populations are harder to separate on the categorical ASUDAS scale than by means of the quantitative maximum depth metric. Still, no Chinese specimen was visually rated below grade 2, and very few European and African specimens were recorded as grade 2 or above, which indicates that, despite the subjective variability, the samples can usually be dichotomized by human raters into high or low shoveling with reasonable accuracy.

When comparing the correlation between the visual assessments (based on the ASUDAS definition and plaque) and some objectively obtained measurements (e.g., predicted ASUDAS values, maximum depth, hollow area), we found that the maximum depth and hollow area metrics correlated highly with visual scoring, even to a higher level than with the predicted ASUDAS values (Table 4). This could suggest that there is a major unconscious reaction to the maximum depth aspect when recording shoveling by following the classical ASUDAS method. In order to test whether the observers were more influenced visually by the depth of the palatal aspect than by the global morphology, we dichotomized the observer ratings following the standard protocol for ASUDAS traits (Scott and Irish, 2017, Scott et al., 2018), which splits the categories of this trait into two broad groups (grades ≤ 1 vs. grades ≥ 2), and re-ran ICC consistency and accuracy measures (Table 5). This grouping mimicked the distinction between East Asians and the rest by separating absent/low degrees of shoveling and marked shoveling. The new ICC values showed high degrees of precision and accuracy, suggesting that the visual scoring performed by observers was largely successful at separating low degrees of shoveling from marked shoveling, but not so successful in detecting finer differences between the ASUDAS casts. This result is relevant as it shows that, while the ASUDAS method is efficient to distinguish below and above the breakpoint grade 2, it is more prone to bias when dealing with close grades and limits the possibilities for more advanced analyses. For example, an important question regarding dental traits is to find which genetic factors are responsible for a trait's expression. Scott and Turner (1997) used the example of incisor shoveling to conclude that such nonmetric dental traits possibly arise from

an underlying continuous trait which is likely polygenic. They also note that dichotomizing such traits leads to the loss of a large amount of variation in the trait. As an extreme example, all Native Americans may have the constant value of 'present' for the shoveling trait, which causes it to "lose its status as a nonmetric variant as it is present in all individuals." Such loss of variation, either when constructing a 'quasi continuous' (ordinal) trait from a continuous underlying trait, or when dichotomizing an ordinal trait, reduces the resolution of the data. This loss of variation leads to a loss of power in genetic association analysis, where the use of a continuous trait can entail 'significantly higher power', especially with small sample sizes (Bhandari et al., 2002). In their literature review, Scott et al. (2018) also note that such simplifications can create several problems for genetic analyses, e.g. the simplified traits can "sometimes mimic the segregation patterns of simple Mendelian inheritance where, in reality, inheritance is complex." (Scott et al., 2018:133). Even in the context of assessing rater reliability, the higher resolution offered by continuous data provides much better reliability estimates (Donner and Eliasziw, 1994).

In the context of clinical studies, Altman (2006) comments that dichotomized variables often appear to be more alluring as they simplify the data while retaining the main dichotomy that is thought to be the crux of the variable, thereby leading to simpler interpretations as well as higher rater agreement. Yet such deliberate discarding of data causes several problems: loss of power, increased risk of false positives, underestimation of variation within or between groups, loss of information about the relationship between the trait and other variables, and increased confounding with other variables in regression analysis, such as genetic association analysis. MacCallum et al. (2002) report similar criticisms, and emphasize the problems in statistical analysis. Using a dichotomous variable means that many statistical procedures are not applicable, e.g., in a genetic association study, the standard linear regression model, which allows estimation of effect sizes in absolute units, cannot be used. Dichotomous variables can only be used with logistic regression, which only estimates effect sizes on an odds ratio scale, making it much more difficult to interpret the effect of a genetic marker or combine evidence from multiple studies.

The sample-dependent nature of Principal Components means that the PCs depend on the whole sample composition. We investigated how the PCs could change when the dataset was limited to European samples, which span only a narrow range of the trait variability (without any marked shoveling). To assess this, we performed the alignment and PC calculation steps for both Procrustes and non-Procrustes methods while restricting the dataset to French samples (contemporary and medieval), and obtained their correlation with the directly

measured metrics. For these correlation values, the correlation of PC1 from both analyses with maximum depth decreased a little (compared to Table 3), especially for the Procrustes PC (Supporting Information Table 4). This indicates that PCs calculated in this reduced dataset with a narrower range of variability might have some increased noise or may be slightly less able to capture the actual underlying metric of lower grades. To verify this, we looked at the correlation of rater-assigned ASUDAS categories of the French samples with these PCs and directly measured metrics (Supporting Information Table 5). In contrast to Table 4, correlations of the rater-assigned ASUDAS categories with maximum depth are higher than their correlations with PC1 by a larger margin.

Discussion

In the literature, the significance of errors due to the role of the observer in the visual scoring of dental non-metric traits using the ASUDAS method is usually considered as being low and/or negligible (Scott and Turner, 1997; Bailey and Hublin, 2013; Hillson, 2005). Applied to decades of research, the classical ASUDAS method has proved to be quite efficient for inferring biological relationships among modern humans (Scott and Turner, 1997; Scott et al., 2018), living non-human primates (Pilbrow, 2003) and fossil hominins (Mizoguchi, 1985; Crummet, 1994; Bailey and Hublin, 2013; Irish et al., 2013; Bailey and Wood, 2007; Martinon-Torres et al., 2007) – in part because, once the grades are scored, they are dichotomized into presence/absence to help reduce observer error and because of the current dichotomous biological distance statistics available (e.g., MMD). There has been a previous attempt to link morphology and measurements for the UI1 shoveling trait, notably by considering the depth of the lingual fossa with respect to ASUDAS grades (Hanihara, 2008). However, this method only considers the maximum depth at the center of the lingual fossa and does not quantify the shape of this fossa. Thus, it is still possible to develop innovative, complementary methods. The recent development of quick and efficient methods for acquiring 3D models of an object (e.g., photogrammetry, laser scanner), together with the advent of powerful quantitative techniques to assess shape variation (geometric morphometrics), has opened up new ways to test the reliability (precision and accuracy) of the ASUDAS method. These methods represent an opportunity to provide objective protocols to investigate non-metric dental variation. In this preliminary study, we have compared the classical plaque-based visual scoring assessment with a new 3D geometric morphometric approach. We propose here a simple, fast method based on geometric morphometrics to characterize a sample of modern human UI1s using a continuous scale of morphological

variation of shoveling. The intra-observer error related to the visual scoring is very low, as previously demonstrated for ASUDAS plaques (Nichol and Turner, 1986; Scott and Turner, 1997; Scott, 2008; Scott and Irish, 2017). As anticipated, our results highlight some limitations of the use of the current ASUDAS plaque, indicating that it did not necessarily represent an objective gradient of expression of a non-metric tooth feature (Figures 3, 4 and 6). Our results also agree with the currently recognized ASUDAS breakpoint between the recorded absence (grades 0-1) and presence (grades 2-7) of shoveling (Scott and Irish, 2017, Scott et al., 2018). This method can also distinguish between the French and South African groups (expressing low degrees of shoveling) and the Chinese sample (being more variable but mostly showing well-defined shovel-shaped incisors). This is in agreement with the vast literature on the topic (e.g., Scott and Turner, 1997; Irish and Scott, 2016; Scott and Irish, 2017; Scott et al., 2018) and demonstrates that our method, while confirming the ASUDAS results, opens a path towards more advanced quantitatively-based assessment for the distinction of fossil and extant human populations. This modest sample was only used here to test the method, but by increasing it and incorporating larger chrono-geographic groups, including fossil hominins, there is a high potential to better understand the evolution of shovel-shaped incisors. For example, Neanderthals are well-known for their markedly shovel-shaped incisors and, given the increasing availability of 3D virtual data on their teeth, paleogenetics techniques and molecular data on tooth morphology (Zanolli et al., 2017), this new quantitative method is perfectly suited to the investigation of the evolution of UI1 shoveling.

Our protocol integrates the analysis of two different but complementary aspects: the depth of the lingual surface with respect to the marginal ridges and the shape of the lingual aspect. This is an important point as our analyses have revealed that visual rating of shoveling tends to be more prone to intra- and inter-observer bias for the highest grades (even starting at grade 2). In addition, even when the observers are well trained and follow the definition of the UI1 shoveling trait (Supporting Information Table 2), when dealing with numerous specimens, they tend to create a mental image of the ASUDAS categories and then make their judgments, resulting in a mental scale that is linearly dependent on the maximum depth of the palatal aspect, while the ASUDAS grades are not distributed linearly for this parameter. This results in the visually assigned ratings being correlated with the maximum depth rather than with predicted ASUDAS categories. In this context, our results provide a reliable, reproducible framework reinforced by statistical results supporting the fact that open scale numerical measurements can complement the ASUDAS method and provide new information. Of

course, similar methods complementing the classic ASUDAS method still need to be developed for other non-metric dental traits. There are also other possibilities for the quantitative study of shape variation, with or without landmarks. For example, a surface deformation-based approach considering a 3D portion of the crown surface (such as the lingual aspect in the case of UI1 shoveling) could be used to assess the degree of deformation from one tooth to another and quantify shape variations of the complete set of dental traits (Durrleman et al.,2012; Durrleman et al., 2014). Thus, although the ASUDAS is a reliable and efficient tool, it is still possible to complement it with alternative methods.

Acknowledgements

This study is based on the PhD research program of the first author. It is supported by the French CNRS. The work of P. Monsarrat is supported by Toulouse University Hospital (CHU de Toulouse), Université Toulouse III - Paul Sabatier, the Midi-Pyrenées region and the research platform of the Toulouse Dental Faculty (PLTRO). J. Braga provided access to the ASUDAS plaque. We thank G. Krüger for granting access to the Pretoria Bone Collection (PBC) used in this study. For scientific discussion, we are also grateful to F. Duret.

References

Altman, D.G. (2006). The cost of dichotomising continuous variables. *The British Medical Journal*, **332**, 1080.

Bailey, S.E., & Wood, B.A. (2007). Trends in postcanine occlusal morphology within the hominin clade: the case of *Paranthropus*. In S.E. Bayley & J.-J. Hublin (Eds.), *Dental Perspectives on Human Evolution: State of the Art Research in Dental Anthropology* (pp. 33-52). Dordrecht: Springer.

Bailey, S.E., & Hublin J-J. (2013). What does it mean to be dentally "modern"? In G.R. Scott & J.D. Irish (Eds.), *Anthropological Perspectives on Tooth Morphology, Genetics, Evolution, Variation* (pp. 222-249). Cambridge: Cambridge University Press.

Bhandari, M., Lochner, H., & Tornetta P. (2002). Effect of continuous versus dichotomous outcome variables on study power when sample sizes of orthopaedic randomized trials are small. *Archives of Orthopaedic and Trauma Surgery*, **122**, 96-98.

Carbonell, V.M. (1963). Variations in the frequency of shovel-shaped incisors in different populations. In D.R. Brothwell (Ed.), *Dental Anthropology* (pp. 211-234). New York: Pergamon Press.

Carter, K., Worthington, S., & Smith, T. (2014). Non-metric dental traits and hominin phylogeny. *Journal of Human Evolution,* **69**, 123-128.

Coleman, M.N., & Colbert, M.W. (2007). CT thresholding protocols for taking measurements on three-dimensional models. *American Journal of Physical Anthropology*, **133**, 723-725.

Crummett, T. (1994). *The evolution of shovel shaping: regional and temporal variation in human incisor morphology*. Ph.D. Dissertation, University of Michigan, Ann Arbor.

Crummett, T. (1995). The three dimensions of shovel-shaping. In J. Moggi-Cecchi (Ed.), *Aspects of Dental Biology: Palaeontology, Anthropology and Evolution* (pp. 305-313). Florence: International Institute for the Study of Man.

Dahlberg, A.A. (1956). *Materials for the Establishment of Standards for Classification of Tooth Characteristics, Attributes, and Techniques in Morphological Studies of the Dentition*. Chicago: Zoller Laboratory of Dental Anthropology, University of Chicago.

Dahlberg, A.A. & Mikkelsen, O. (1947). The shovel-shaped character in the teeth of the Pima Indians. *American Journal of Physical Anthropology*, **5**, 234-235.

Denton, L.C. (2011). *Shovel-Shaped Incisors and the Morphology of the Enamel-Dentin Junction: An Analysis of Human Upper Incisors in Three Dimensions*. Master dissertation, Colorado State University, Fort Collins.

Donner, A., & Eliasziw, M. (1994). Statistical implications of the choice between a dichotomous or continuous trait in studies of interobserver agreement. *Biometrics*, **50**, 550-555.

Dray, S. & Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1-20.

Dryden, I.L. (2017). shapes package. R Foundation for Statistical Computing, Vienna, Austria. Contributed package. Version 1.2.3. URL http://www.R-project.org.

Durrleman, S., Prastawa, M., Korenberg, J.R., Joshi, S., Trouvé, A., & Gerig, G. (2012). Topology preserving atlas construction from shape data without correspondence using sparse parameters. In N. Ayache, H. Delingette, P. Golland & K. Mori (Eds.), *Proceedings of Medical Image Computing and Computer Aided Intervention* (pp. 223-230). Nice: Springer.

Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., & Trouvé, A. (2014). Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage*, **101**, 35-49.

Fajardo, R.J., Ryan, T.M., & Kappelman J. (2002). Assessing the accuracy of high-resolution X-ray computed tomography of primate trabecular bone by comparisons with histological sections. *American Journal of Physical Anthropology*, **118**, 1-10.

Gamer, M., Lemon, J., & Puspendra Singh, I. (2012). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84. https://CRAN.R-project.org/package=irr.

Goaz, P.W., & Miller, M.C. (1966). A preliminary description of the dental morphology of the Peruvian Indians. *Journal of Dental Research*, **45**, 106-119.

Hanihara, K. (1969). Mongoloid dental complex in the permanent dentition. *Proceedings of the VIIIth International Congress of Anthropological and Ethnological Sciences*, **1**, 298-300.

Hanihara, K., Masuda, T., Tanaka, T. (1975). Family studies of the shovel trait in the maxillary central incisor. *Journal of the Anthropological Society of Nippon*, **83**,107-12.

Hanihara, K. (2008). Morphological variation of major human populations based on nonmetric dental traits. *American Journal of Physical Anthropology*, **136**, 169-182.

Hillson, S.W. (1996). Dental anthropology. Cambridge: Cambridge University Press.

Hillson, S.W. (2005). Teeth. Cambridge: Cambridge University Press.

Hlusko, L.J., Carlson, J.P., Chaplin, G., Elias, S.A., Hoffecker, J.F., Huffman, M., Jablonski, N.G., Monson, T.A., O'Rourke, D.H., Pilloud, M.A., & Scott, G.R. (2018). Environmental selection during the last ice age on the mother-to-infant transmission of vitamin D and fatty acids through breast milk. *Proceedings of the National Academy of Sciences of the United States of America* (in press, doi.org/10.1073/pnas.1711788115).

Hoffman, J.W., & de Beer, F.C. (2012) Characteristics of the Micro-Focus X-ray Tomography Facility (MIXRAD) at Necsa in South Africa. In *Proceedings of the 18th World Conference on Nondestructive Testing* (pp. 1-12). Durban: South African Institute for Non-Destructive Testing.

Hrdlička, A. (1920). Shovel-shaped teeth. *American Journal of Physical Anthropology*, **3**, 429-465.

Hughes, I., & Hase, T. (2010). *Measurements and Their Uncertainties: A Practical Guide to Modern Error Analysis*. Oxford: Oxford University Press.

Irish, J.D. (1998). Ancestral dental traits in recent Sub-Saharan Africans and the origins of modern humans. *Journal of Human Evolution*, **34**, 81-98.

Irish, J.D., & Guatelli-Steinberg, D. (2003). Ancient teeth and modern human origins: An expanded comparison of African Plio-Pleistocene and recent world dental samples. *Journal of Human Evolution,* **45**, 113-144.

Irish, J.D., Guatelli-Steinberg, D., Legge, S.S., Ruiter, D.J., & Berger, L.R. (2013). Dental morphology and the phylogenetic "place" of *Australopithecus sediba. Science*, **340**, 1233062.

Irish, J.D. (2014). Dental nonmetric variation around the world: Using key traits in populations to estimate ancestry in individuals. In G.E. Berg & S.C. Ta'ala (Eds.), *Biological Affinity in Forensic Identification of Human Skeletal Remains: Beyond Black and White* (pp. 165-190). Boca Raton: CRC Press-Taylor and Francis Group.

Irish, J.D., Guatelli-Steinberg, D., Legge, S.S., Ruiter, D.J., & Berger, L.R. (2014). Response to "Non-metric dental traits and hominin phylogeny" by Carter et al., with additional information on the Arizona State University Dental Anthropology System and phylogenetic "place" of *Australopithecus sediba*. *Journal of Human Evolution,* **69**, 129-134.

Irish, J.D., & Scott, G.R. (2016). *A Companion to Dental Anthropology*. Chichester: Wiley Blackwell.

Jernvall, J., Keränen, S.V.E., & Thesleff, I. (2000). Evolutionary modification of development in mammalian teeth: Quantifying gene expression patterns and topography. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 14444–14448.

Joint Committee for Guides in Metrology. (2008). *The International Vocabulary of Metrology, Basic and General Concepts and Associated Terms*.

Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K., Hanihara, T., Matsukusa, H., Kawamura, S., Maki, K., Osawa, M., Ishida, H., & Oota, H. (2009). A common variation in EDAR is a genetic determinant of shovel shaped incisors. *The American Journal of Human Genetics*, **85**, 528-535.

Koo, T.K., Li, M.Y. (2017). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, **15**, 155-163.

Le Cabec, A., Gunz, P., Kupczik, K., Braga, J., & Hubin, J.J. (2013). Anterior tooth root morphology and size in Neanderthals: taxonomic and functional implications. *Journal of Human Evolution,* **64**, 169-193.

Lehnert, B. (2015). BlandAltmanLeh package. R Foundation for statistical computing, Vienna, Austria. Contributed package. Version 0.3.1. Available from: http://www.R-project.org.

Ligges, U., & Mächler, M. (2003). Scatterplot3d - an R Package for Visualizing Multivariate Data. *Journal of Statistical Software*, **8**, 1-20.

MacCallum, R.C., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, **7**, 19-40.

Martinón-Torres, M., Bermúdez de Castro, J.M., Gómez-Robles, A., Arsuaga, J.L., Carbonell, E., Lordkipanidze, D., Manzi, G., & Margvelashvili, A. (2007). Dental evidence on the hominin dispersals during the Pleistocene. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 13279-13282.

MATLAB and Statistics Toolbox Release. (R2017ba). The MathWorks, Inc., Natick, Massachusetts, United States.

Mizogushi, Y. (1985). Shoveling: a statistical analysis of its morphology. *The University Museum, the University of Tokyo Bulletin*, **26** .

Mizogushi, Y. (2013). Significant among-population associations found between dental characters and environmental factors. In G.R. Scott & J.D. Irish (Eds*.), Anthropological Perspectives on Tooth Morphology, Genetics, Evolution, Variation* (pp. 108-125). Cambridge: Cambridge University Press.

Nichol, C.R., & Turner, C.G., II (1986). Intra- and interobserver concordance in classifying dental morphology. *American Journal of Physical Anthropology*, **69**, 299-315.

Park, J.H., Yamaguchi, T., Watanabe, C., Kawaguchi, A., Haneji, K., Takeda, M., Kim, Y.I., Tomoyasu, Y., Watanabe, M., Oota, H., Hanihara, T., Ishida, H., Maki, K., Park, A.B., & Kimura, R. (2012). Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *Journal of Human Genetics*, **57**, 508-514.

Pilbrow, V. (2003). *Dental variation in African apes with implications for understanding patterns of variation in species of fossil apes*. PhD Dissertation. New York: University of New York.

R Core Team. (2017). R Foundation for statistical computing. Vienna, Austria. Version 3.4.3. Available from: http://www.R-project.org.

Salazar-Ciudad, I. & Jernvall, J. (2002). A gene network model accounting for development and evolution of mammalian teeth. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 8116-8120.

Salazar-Ciudad, I. & Jernvall, J. (2010). A computational model of teeth and the developmental origins of morphological variation. *Nature*, **464**, 583-586.

Scott, G.R (1973). *Dental Morphology: a genetic study of American White Families and Variation in living Southwest Indians*. PhD dissertation, Department of Anthropology, Arizona State University, Tempe.

Scott, G.R. (2008). Dental Morphology. In A. Katzenburg & S. Saunders (Eds.), *Biological Anthropology of the Human Skeleton, 2nd Edition* (pp. 265-298). New York: Wiley-Liss.

Scott, G.R., & Irish, J.D. (2017). *Human Tooth Crown and Root Morphology. The Arizona State University Dental Anthropology System*. Cambridge: Cambridge University Press.

Scott, G.R., & Turner, C.G. (1997). *The Anthropology of Modern Human Teeth: Dental Morphology and its Variation in Recent Human Populations, 2nd Edition*. Cambridge: Cambridge University Press.

Scott, G.R., Turner, C.G., Townsend, G.C., & Martinón-Torres, M. (2018). *The Anthropology of Modern Human Teeth. Dental Morphology and Its Variation in Recent and Fossil* Homo sapiens. Cambridge: Cambridge University Press.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420-428

Smith, H. B. (1984). Patterns of molar wear in hunter-gatherers and agriculturalists. *American Journal of Physical Anthropology*, **63**, 39-56.

Spoor, C.F., Zonneveld, F.W., & Macho, G.A. (1993). Linear measurements of cortical bone and dental enamel by computed tomography: Applications and problems. *American Journal of Physical Anthropology*, **91**, 469-484.

Taverne, P.P., Amesz-Voorhoeve, W.H.M., Leertouwer, H.L. (1979). A photogrammetric method which provides geometric data in dental morphology. *Zeitschrift fur Morphologie und Anthropologie*, **70**, 163-173.

Turner, II C.G. (1990). Major features of sundadonty and sinodonty, including suggestions about East Asian microevolution, population history, and Late Pleistocene relationships with Australian aboriginals. *American Journal of Physical Anthropology*, **82**, 295-317.

Turner, II C.G., Nichol, C.R., & Scott, G.R. (1991). Scoring procedures for key morphological traits of the permanent dentition: the Arizona State University Dental Anthropology System. In M.A. Kelley & C.S. Larsen (Eds.), *Advances in Dental Anthropology* (pp. 13-31). New York, Wiley-Liss.

Zanolli, C. (2013). Additional evidence for morpho-dimensional tooth crown variation in a new Indonesian *H. erectus* sample from the Sangiran Dome (Central Java). *PLoS One*, **8**, e67233.

Zanolli, C., Bondioli, L., Coppa, A., Dean, M.C., Bayle, P., Candilio, F., Capuani, S., Dreossi, D., Fiore, I., Frayer, D.W., Libsekal, Y., Mancini, L., Rook, L., Medin, T., Tuniz, C., & Macchiarelli, R. (2014). The late Early Pleistocene human dental remains from Uadi Aalad and Mulhuli-Amo (Buia), Eritrean Danakil: macromorphology and microstructure. *Journal of Human Evolution*, **74**, 96-113.

Zanolli, C., Hourset, M., Esclassan, R., & Mollereau, C. (2017). Neanderthal and Denisova tooth protein variants in present-day humans. *PLoS One*, **12**, e0183802.

Captions for figures

Figure 1. (a). Position of the reference plane located at the midpoint of the crown, which was used to place the semilandmarks along the curve of the lingual aspect of the crown from the mesial to the distal side (b). The maxima of the extreme curvature line were used as starting and ending points of our GM analysis

Figure 2. Illustration of the maximum depth and hollow area (a) used in the non-Procrustes analyses (semilandmark curves aligned with their first and last points). The ASUDAS reference plaque teeth curves are superimposed following this non-Procrustes approach, showing the non-linear variation in shape from grade 0 to grade 6 (b).

Figure 3. Principal component analysis (PCA) of the Procrustes-registered shape coordinates of the 100 semilandmarks used as proxy to assess UI1 shoveling. A: PC1 vs. PC2; B: PC3 vs. PC4.

Figure 4. Histograms showing the frequency of Procrustes PC1 (a), maximum depth (b) and hollow area (c) values for the 87 modern human specimens and the distribution of the ASUDAS reference grades (black vertical lines).

Figure 5. Bland-Altman visualization for agreement of the ASUDAS visual observations and Procrustes coordinates of the semilandmark curves. The different agreements between the two raters were plotted for X, Y and Z float coordinates. Among the 100 landmarks, only the first, mid and last landmarks are drawn.

Figure 6. Plots of the maximum depth against the visual ASUDAS scoring of the first observer's tests (VS1 T1: a; VS1 T2: b) and the second rater (VS2: c). The black dots correspond to the values of the ASUDAS reference plaque and they were joined up via spline interpolation. The symbols represent the chrono-geographic origin as indicated in the legend of the graphs.

Table 1. List of 87 modern human UI1 elements considered in this study.

Time period	Geographic origin	depository
contemporary (n=13)	France	MHNT[1]
medieval(n=31)	France	INRAP[2]
contemporary (n=13)	China	IVPP[3]
contemporary (n=30)	South Africa	PBC[4]

[1]MHNT: Natural History Museum of Toulouse; [2]INRAP: French National Institute for Preventive Archeological Research; [3]IVPP: Institute of Vertebrate Paleontology and Paleoanthropology of Beijing; [4]PBC: Pretoria Bone Collection of the Department of Anatomy of the University of Pretoria, representing individuals of various ancestries including Nbele, N Sotho, Swazi and Zulu.

Table 2. Intra-Class Correlation (ICC) values for intra-observer (VS1 T1 vs. T2) and inter-observer (VS1 vs. VS2) consistency measures and accuracy (comparing the three ratings with their predicted values from the ASUDAS reference plate teeth). Shades of color represent the degree of consistency and accuracy, with darker green corresponding to the highest degrees and white to the lowest degree.

	consistency (precision)		accuracy (unbiasedness)		
	VS1 T1 vs T2	VS1 vs VS2	VS1 T1 vs predicted	VS1 T2 vs predicted	VS2 vs predicted
all samples	0.985	0.987	0.759	0.749	0.772
South African	0.946	0.986	0.602	0.565	0.602
French contemporary	0.900	0.721	0.496	0.570	0.494
French medieval	0.971	0.993	0.550	0.592	0.559
Chinese	0.976	0.975	0.000	0.000	0.000

Table 3. Correlations between various measurements obtained from the coordinates of all samples (87 human specimens plus the 7 ASUDAS reference casts). Maximum depth and hollow area are the directly measured metrics based on the non-Procrustes alignment method. Principal components were obtained from both Procrustes and non-Procrustes methods. Shades of color represent the degree of correlation, with darker green corresponding to the highest degrees and white to the lowest degree. Red indicates negative correlation.

		direct metrics		**Procrustes PC**		**non-Procrustes PC**	
		maximum depth	hollow area	PC1	PC2	PC1	PC2
direct metrics	maximum depth	-	0.98	0.98	0.08	0.99	-0.08
	hollow area	0.98	-	0.96	0.20	1.00	0.06
Procrustes PC	PC1	0.98	0.96	-	0.00	0.97	-0.21
	PC2	0.08	0.20	0.00	-	0.16	0.86
non-Procrustes PC	PC1	0.99	1.00	0.97	0.16	-	0.00
	PC2	-0.08	0.06	-0.21	0.86	0.00	-

Table 4. Correlation of observer-assigned ASUDAS categories with the predicted ASUDAS value, the first four PCs of the Procrustes analysis, the maximum palatal depth and the hollow area. Shades of color represent the degree of correlation, with darker green corresponding to the highest degrees and white to the lowest degree. Red indicates negative correlation.

	VS1 T1	VS1 T2	VS2
predicted ASUDAS	0.776	0.769	0.788
PC1	0.836	0.831	0.850
PC2	0.153	0.169	0.177
PC3	-0.106	-0.097	-0.082
PC4	0.055	0.028	0.056
maximum depth	0.840	0.833	0.852
hollow area	0.832	0.828	0.848

Table 5. ICC values for intra-observer (VS1 T1 vs. T2) and inter-observer (VS1 vs. VS2) consistency measures and accuracy (comparing the three ratings with their predicted values from the ASUDAS reference plate teeth) when grades 0-1 and 2-6 are fused into two groups (2-category split). Shades of color represent the degree of consistency and accuracy, with darker green corresponding to the highest degrees and white to the lowest degree.

	consistency (precision)		accuracy (unbiasedness)		
	VS1 T1 vs T2	VS1 vs VS2	VS1 T1 vs predicted	VS1 T2 vs predicted	VS2 vs predicted
all samples	0.945	0.964	0.812	0.827	0.842

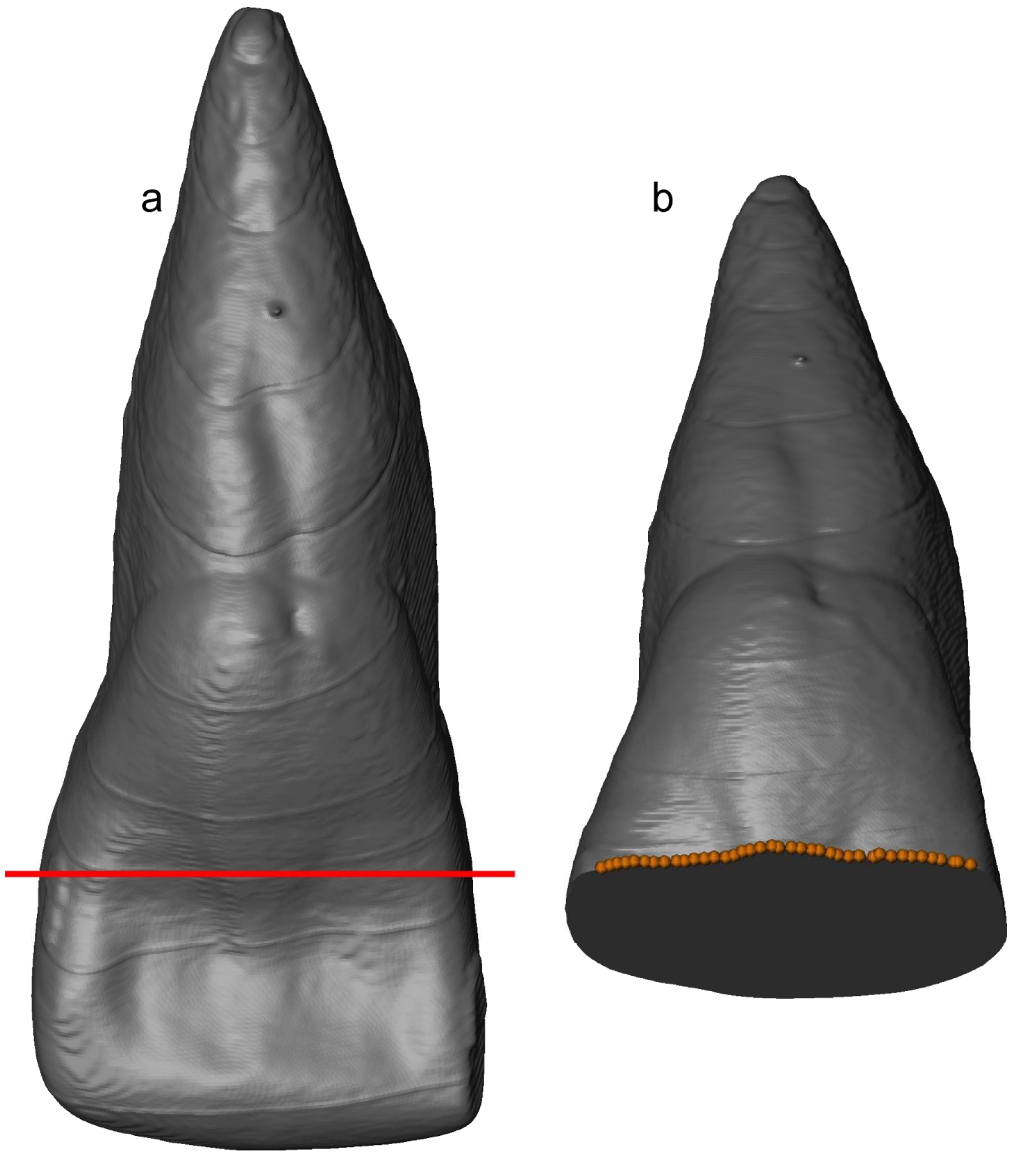

Figure1.(a).Position of the reference plane located at the midpoint of the crown, that is used to put the semilandmarks along the curve of the lingual aspect of the crown from mesial to distal side. (b).We use the maxima of the extreme curvature line as starting and ending points of our GM analysis.

356x407mm (300 x 300 DPI)

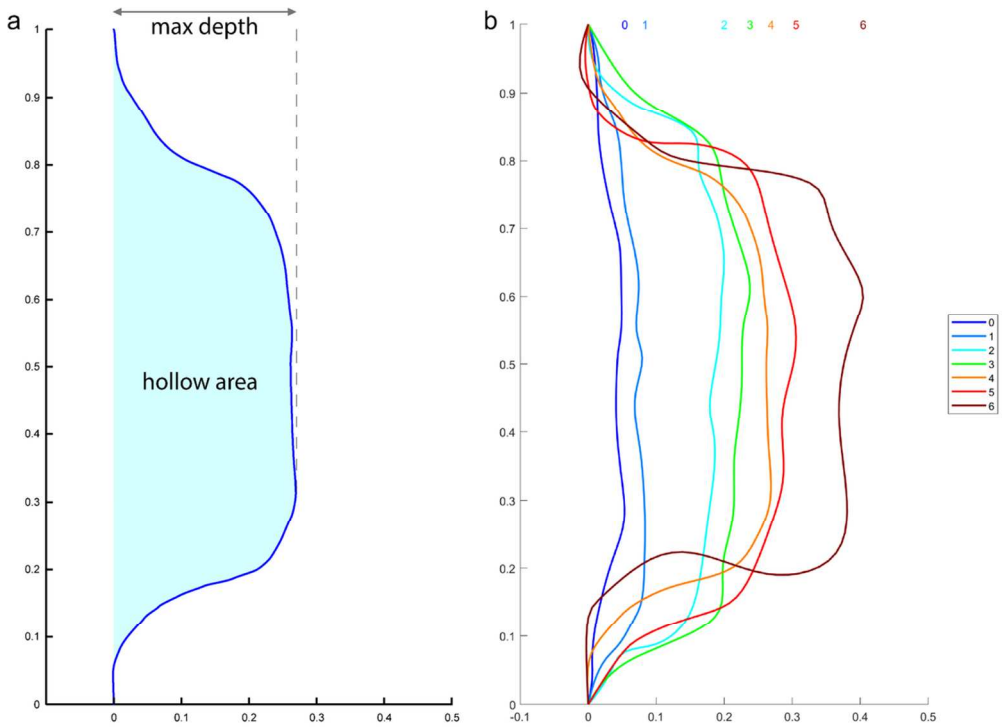

Figure 2. Illustration of the maximal depth and hollow area (a) used in the non-Procrustes analyses (semilandmark curves aligned with their first and last points). The ASUDAS reference plaque teeth curves are superimposed following this non-Procrustes approach, showing the non-linear variation in shape from grade 0 to grade 6 (b).

106x76mm (300 x 300 DPI)

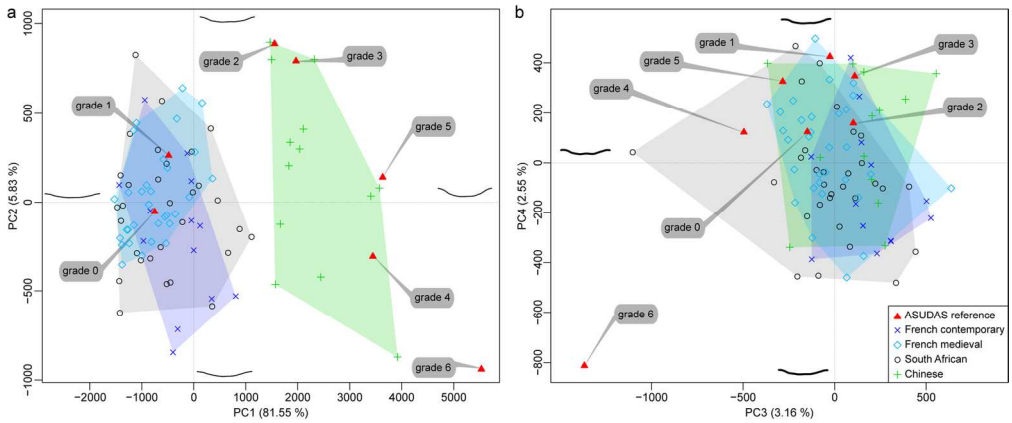

Figure 3. Principal component analysis (PCA) of the Procrustes-registered shape coordinates of the 100 semilandmarks used as proxy to assess UI1 shoveling. A: PC1 vs. PC2; B: PC3 vs. PC4.

154x64mm (300 x 300 DPI)

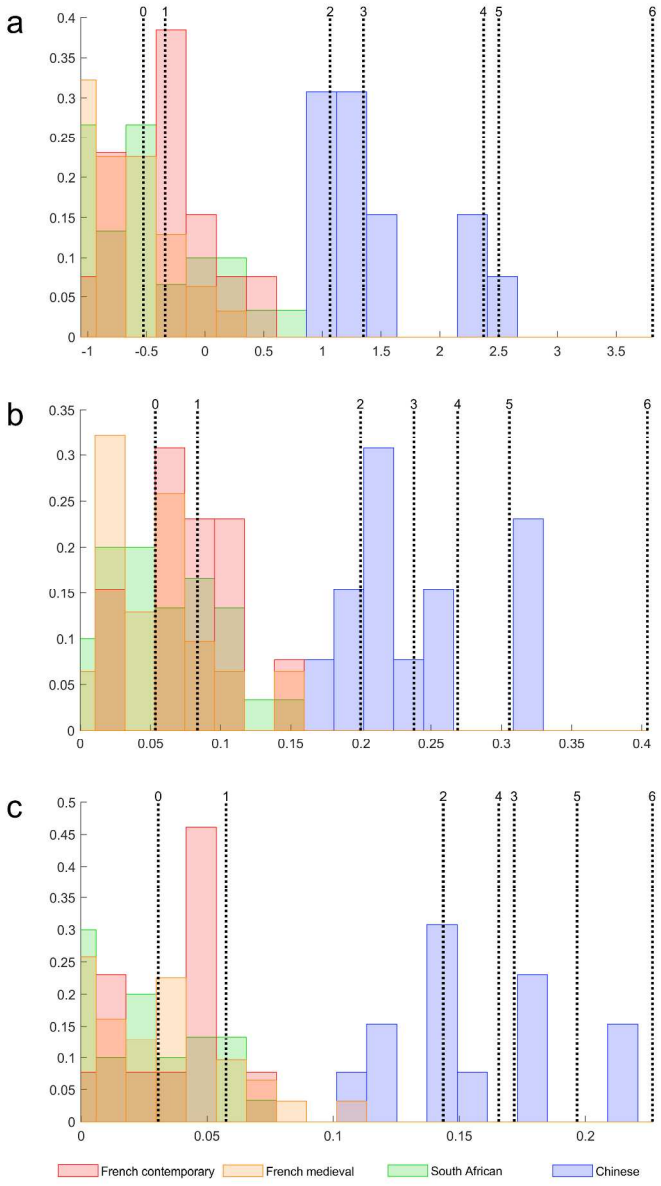

Figure 4. Histograms showing the frequency of Procrustes PC1 (a), maximum depth (b) and hollow area (c) values for the 87 modern human specimens and the distribution of the ASUDAS reference grades (black vertical lines).

187x319mm (300 x 300 DPI)

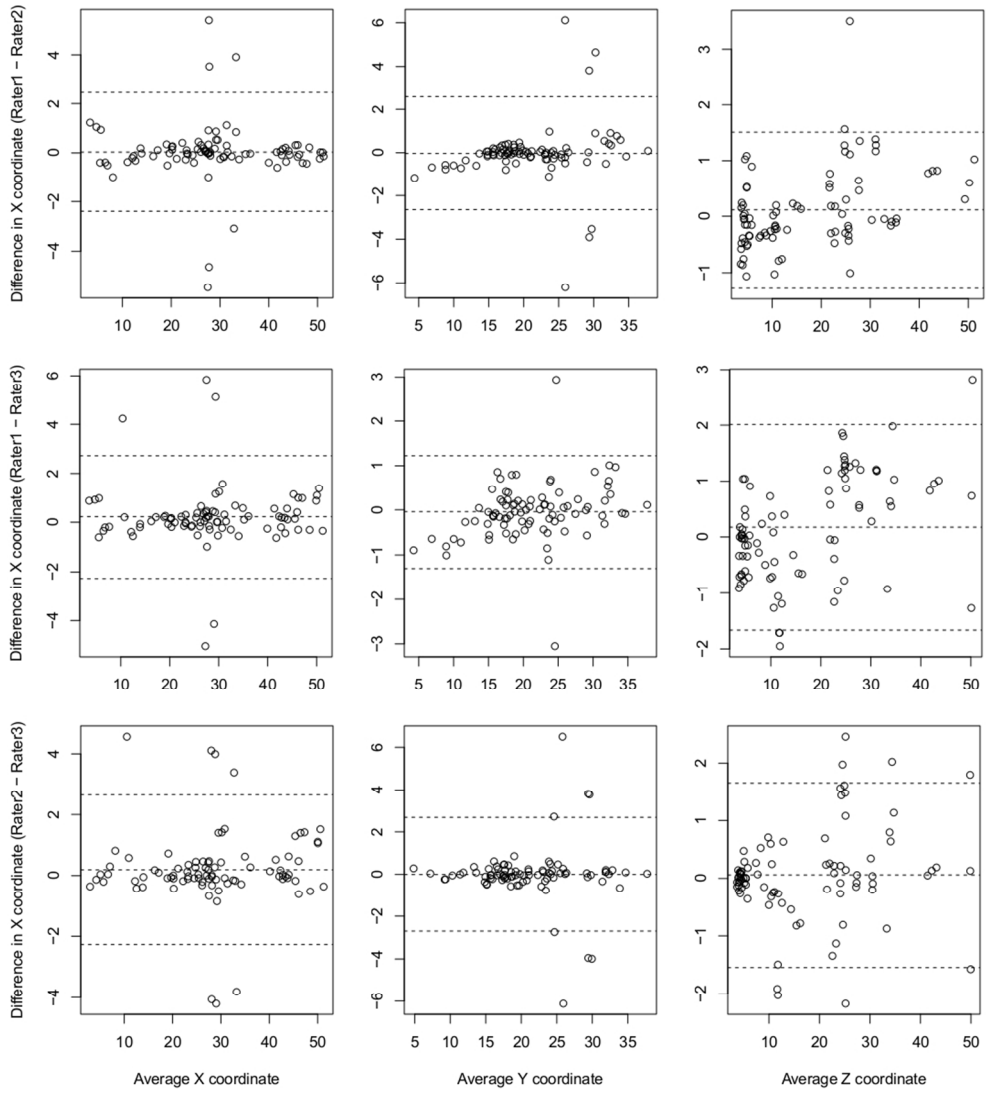

Figure 5. Bland-Altman visualization for agreement of the ASUDAS visual observations and Procrustes coordinates of the semilandmark curves. The different agreements among the pair of raters were plotted for X, Y and Z float coordinates. Among the 100 landmarks, only first, mid and last landmarks were drawn.

92x103mm (300 x 300 DPI)

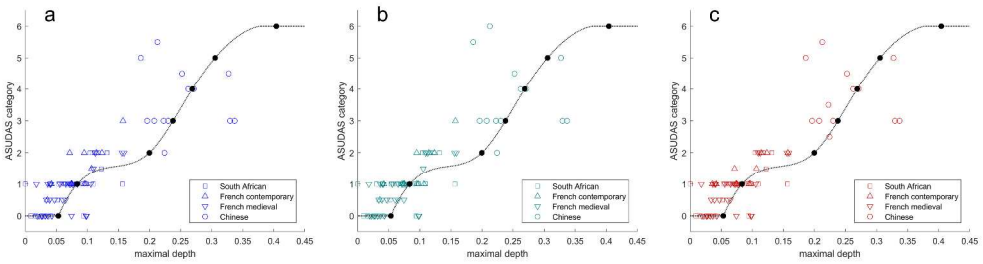

Figure 6. Plots of the maximum depth against the visual ASUDAS scoring of the first observer's tests (VS1 T1: a; VS1 T2: b) and the second rater (VS2: c). The black dots correspond to the values of the ASUDAS reference plaque and they joined between them via spline interpolation. The symbols represent the chrono-geographic origin as indicated in the legend of the graphs.

411x111mm (300 x 300 DPI)