

The long and winding road: identifying pig domestication through molar size and shape

Allowen Evin^{a,b,*}, Thomas Cucchi^{a,b}, Andrea Cardini^{c,d,e}, Una Strand Vidarsdottir^f, Greger Larson^g, Keith Dobney^a

^a Department of Archaeology, University of Aberdeen, St. Mary's Building, Elphinstone Road, Aberdeen AB24 3UF, United Kingdom

^b CNRS-Muséum National d'Histoire Naturelle, UMR 7209, Archéozoologie, histoire des sociétés humaines et des peuplements animaux, 55 rue Buffon, 75005 Paris, France

^c Dipartimento di Biologia, Università di Modena e Reggio Emilia, via Campi 213, 41100 Modena, Italy

^d Centre for Anatomical and Human Sciences, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom

^e Centre for Forensic Science, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

^f Department of Anthropology, Durham University, South Road, Durham DH1 3LE, United Kingdom

^g Durham Evolution and Ancient DNA, Department of Archaeology, Durham University, South Road, Durham DH1 3LE, United Kingdom

ARTICLE INFO

Article history:

Received 13 June 2012

Received in revised form

1 August 2012

Accepted 4 August 2012

Keywords:

Geometric morphometrics

Randomization

Classification

Sus scrofa

Predictive discriminant analysis

Teeth

Zoarchaeology

ABSTRACT

The ability to document the effects of domestication from archaeological remains of animals and plants is essential for reconstructing the history of one of the most important transitions in human prehistory – the shift from hunting and gathering to farming. In mammals, teeth are well preserved in archaeological remains and are known to be taxonomically informative. In this study, we compare three sets of dental morphometric descriptors in wild and domestic pigs (*Sus scrofa*) – maximum length, size and shape variables from 2D geometric morphometrics – in order to assess which of the three provides the best ability to correctly distinguish current wild and domestic West Palaearctic pigs. For this purpose, we used predictive linear discriminant analysis with cross-validation taking into account potential bias due to heterogeneous sample sizes and important number of predictors. Classification accuracy of wild and domestic status ranged between 77.3 and 93% depending of the tooth and the descriptor analyzed. However, individual posterior probabilities of correct classification were appreciably smaller when using tooth length and centroid size compared to shape variables. Size appeared to be a poor indicator of wild and domestic status, contrary to shape which in addition provides a high degree of confidence in the wild versus domestic predictions. Our results indicate that geometric morphometrics offers an extremely powerful alternative to more traditional biometric approaches of length and width measurements to capture the elusive morphological changes induced by the domestication process in archaeological remains.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past 10,000 years human subsistence has been transformed by the domestication of plants and animals. Though the differences between 'wild' and 'domestic' animals is generally understood, fundamental questions regarding this basic dichotomy, along with the processes involved with the biological and cultural

transformations leading to domestication, remain largely unanswered (Dobney and Larson, 2006; Vigne et al., 2005). The bones and teeth of animals recovered from archaeological sites can provide direct evidence of this important transition in human history, as they exhibit phenotypic and genotypic changes associated with natural and artificial selection. Wild and domestic forms of the same species are often morphologically, behaviourally and/or ecologically distinct (Price, 2002). In mammals, the most significant physical changes to occur with domestication involve a decrease in brain and body size, changes in body proportions, and a modification of external morphological characters such as emergence of piebald coat colour, wavy or curly hair, rolled and shortened tails, or floppy ears (e.g. Dobney and Larson, 2006; O'Regan and Kitchener, 2005). Notably, a decrease in body size during domestication has

* Corresponding author. Department of Archaeology, University of Aberdeen, St. Mary's Building, Elphinstone Road, Aberdeen AB24 3UF, United Kingdom.

E-mail addresses: a.evin@abdn.ac.uk, evin@mnhn.fr (A. Evin), cucchi@mnhn.fr (T. Cucchi), cardini@unimo.it (A. Cardini), una.vidarsdottir@durham.ac.uk (U. Strand Vidarsdottir), greger.larson@durham.ac.uk (G. Larson), keith.dobney@abdn.ac.uk (K. Dobney).

been demonstrated in many mammals including dogs, cattle, goats, sheep and pigs (e.g. Albarella et al., 2005; Davis, 1981; Hongo et al., 2009; Peters et al., 1999; Zeder et al., 2006).

Identifying domestic forms of *Sus scrofa* is particularly challenging for zooarchaeologists as wild boar are distributed throughout Eurasia (Albarella et al., 2006; Rowley-Conwy et al., 2012). Traditional size measurements (linear distances) of teeth and bones have commonly been used to infer the wild or domestic status of archaeological remains (Vigne et al., 2005). Thus, a reduction in size is generally recognized as one of the primary indicators of the transition from wild to domestic in the archaeological record (e.g. Boessneck and von den Driesch, 1978; Bökönyi, 1974; Meadow, 1989). As a consequence, small individuals are commonly classified as 'domestic' and large ones as 'wild' (see reviews in Albarella et al., 2006; Rowley-Conwy et al., 2012). Although the maximum length of the lower third molar (most frequently used for wild/domestic determinations in zooarchaeological analyses) is generally considered longer in wild pigs (e.g. Rüttimeyer, 1862; Boessneck et al., 1963; Flannery, 1982; Rowley-Conwy et al., 2012; Ervynck et al., 2001), the ranges of molar length measurements show significant overlap between the two groups (Albarella and Payne, 2005; Albarella et al., 2006; Mayer et al., 1998; Payne and Bull, 1988). Despite decades of research, size variation between wild and domestic pigs remains inadequately studied, raising doubts about the accuracy of size measurements to discriminate wild and domestic forms (Mayer et al., 1998). More accurate morphometric methods are, therefore, required to first describe specific morphometric differences between wild and domestic dentitions, and second to determine the best morphometric descriptors to classify specimens more accurately.

Although traditional morphometrics provide important reference data on quantitative variation in morphology, they have inherent limitations such as the difficulty in separating size from shape information, the lack of an effective visualization of analytical results (generally interpreted using summary statistics and coefficients tables) and, more importantly, their inability to accurately preserve the relative positions of the anatomical landmarks between which distances are measured (Adams et al., 2004; Rohlf and Marcus, 1993). To overcome these limitations, developments in morphometric methods during the 1980s and '90s, provided new ways of separating size from shape, to accurately capture the geometric relationships amongst the parts being measured and to visualize analytical results using intuitive shape diagrams (Adams et al., 2004; Rohlf and Marcus, 1993; Bookstein, 1996). This approach known as geometric morphometrics (GMM) has become the mainstream set of techniques used in biological and palaeontological studies and are now increasingly being employed in zooarchaeology to tackle taxonomic issues at the specific and sub-specific level in groups with a complex taxonomy, such as murids (Cucchi et al., 2011a; Valenzuela-Lamas et al., 2011; Cucchi, 2008), equids (Bignon et al., 2005) or cave bear (Seetah et al., 2012). In pigs GMM have already been employed to clarify taxonomy and dispersal in Island South-East Asia (Cucchi et al., 2009; Larson et al., 2007) and the beginnings of its domestication in China (Cucchi et al., 2011b). However, these studies have not dealt specifically with estimating the accuracy of assigning archaeological remains to either wild or domestic forms.

Previous molecular analysis has revealed the likely existence of multiple centres of pig domestication in Eurasia (Larson et al., 2005). In order to exclude phenotypic differences potentially due to independent domestication events this study focuses on modern wild and domestic *S. scrofa* from the West Palaeartic. Analyses were undertaken using lower second (Cucchi et al., 2011b) and lower third (Larson et al., 2007; Cucchi et al., 2009) as well as upper second and third molars.

Classification in morphometrics is commonly achieved using a well established parametric method called linear discriminant function analysis (LDA). LDA looks for linear combinations of variables that maximize differences between predefined groups relative to within group variation. Thus, discriminant functions can be derived using measurements on modern domestic and wild pigs that aid with the classification of archaeological material. LDA is sensitive to assumptions that are easily violated by real data; it is affected by sampling error (small sample size as well as unbalanced designs) and the number of variables used as group predictors; also, LDA tends to over fit the data (i.e., its results are often 'over-optimistic' – Kovarovic et al., 2011). This is why classifications from LDA must be cross-validated and the robustness of their results assessed. This is particularly important in archaeology, where sample sizes are often unbalanced. Furthermore, when using GMM approaches, shape analysis tends to generate numerous variables, whose number may be reduced using ordination techniques such as principal component analysis (PCA) (i.e. Sheets et al., 2006; Baylac and Friess, 2005).

This study aims to: (1) describe the variability of modern wild and domestic West Palaeartic *S. scrofa* using a commonly used traditional measurement (maximum molar length) as well as size and shape variables from geometric morphometrics on lower and upper second and third molars; (2) assess the classification accuracy of the different morphometric descriptors – taking into account how the number of predictors and unequal sample size might affect results; (3) estimate the confidence in the classification accuracy when using the different morphometric descriptors.

This study is the first of a series of analyses designed at developing more definitive standards that could be used to improve the identification of wild and domestic forms in the zooarchaeological record. Accuracy of classification and detailed quantitative descriptions of morphometric differences are fundamental for a better understanding of animal domestication history and the often subtle patterns of variation and covariation produced by human-induced selection over millennia.

2. Material

A total of 972 teeth were analyzed, including 327 upper M2, 163 upper M3, 311 lower M2 and 171 lower M3 (Table 1). The geographic range of wild boar specimens in our samples covers the entire West Palaeartic (i.e. North Africa, Western and Eastern Europe, Russia and Near East, see more details in Supplementary Table 1), whilst domestic pigs represent eleven different modern European breeds (Supplementary Table 1). Due to a complex history of pig domestication, including local domestications and population replacements (Larson et al., 2005, 2007), we focused only on the overall signature of domestication (i.e. differences between wild and domestic pigs). Teeth were measured unilaterally, preferentially from the right side. Third molars erupt later than the second molars (Bull and Payne, 1982), which explains the smaller number of M3 in our dataset. Our samples include adults only (older than 12–14 months), both males and females. Since the sex of archaeological *S. scrofa* specimens is often difficult to

Table 1

Sample sizes for wild and domestic pigs per tooth with the ratio between the numbers of wild boars and domestic pigs.

	UpperM2	UpperM3	LowerM2	LowerM3
Wild	268	123	258	129
Domestic	59	40	53	42
Total	327	163	311	171
Domestic/wild	0.18	0.25	0.17	0.25

assess from fragmented dental rows, males and females in our study were pooled.

3. Methods

3.1. Morphometrics

Data were collected from 2D digital photographs of the occlusal surface of each molar using a standardized protocol (Cucchi et al., 2011a): teeth were positioned with their occlusal view being horizontal, and parallel to the focal plane of the lens using a photographic bubble level. Lateral parallax was controlled by adjusting the two most anterior cusps symmetrically. Pictures were taken using a Reflex Camera (Nikon D90) coupled with a micro length (AF-S Micro Nikkor 60 mm).

As the occlusal surface of pig molars is relatively flat, 2D pictures provide a good approximation of its 3D structure. Landmark and sliding semi-landmark coordinates were measured on the digital images using TpsDig2 v2.16 (Rohlf, 2010a). Semi-landmarks are used to measure outlines or surfaces when landmarks are absent, they require particular caution in the interpretation of results and they are subjected to a specific analytical treatment aimed at mathematically improving the correspondence of points which otherwise lack a simple one-to-one correspondence (Bookstein, 1997; Gunz et al., 2005). The configuration of landmarks and sliding semi-landmarks used is shown in Fig. 1.

Landmark coordinates were superimposed using a Generalized Procrustes Analysis (GPA) (Rohlf and Slice, 1990; Goodall, 1995). During this procedure, all specimens are translated so their centre of gravity (or centroid) of their landmark configuration coincides, normalized to unit centroid size and rotated to minimize the squared summed distances between corresponding landmarks. Centroid size (CS) is the square root of the sum of squared distances of the landmarks from their centroid (Bookstein, 1991), and measures the dispersion of the landmarks around the centroid. CS may be used as a univariate summary of the overall size of the

tooth. In the superimposition, semi-landmarks are allowed to slide along the chord drawn between adjacent points to minimize the sum of Procrustes distances between each individual and the mean shape (Sampson et al., 1996; Perez et al., 2006; Zelditch et al., 2004; Sheets et al., 2004). This procedure allows a reduction of the variance due to the lack of a precise correspondence of the semi-landmarks across the specimens. The superimposition using sliding semi-landmarks was performed in TPS Relw v1.49 (Rohlf, 2010b). Since results obtained for shape and form (size + shape) show very little differences, only results obtained for shape will be presented and discussed.

The new set of shape coordinates obtained from the GPA – namely Procrustes coordinates – together represent the total amount of shape variation in the entire sample. However, these data are partially redundant (i.e. the number of variables is larger than the number of truly informative dimensions). Thus, shape variation was summarized and redundancy removed using a principal component analysis (PCA) on the variance covariance matrix of the Procrustes coordinates. Maximum Tooth Length (MTL) was computed as the longest distance between the most anterior and the most posterior semi-landmarks and is equivalent to the metric traditionally measured with calipers. Size measures (MTL and CS) were log-transformed.

3.2. Statistical analyses

3.2.1. Initial datasets

Distributions of MTL, CS and shape were visualized using frequency histograms. For shape, individual distribution was visualized along the PC1 of a between group principal component analysis (Seetah et al., 2012; and references therein). Differences between wild and domestic specimens in MTL, CS and shape were analyzed using leave-one-out cross-validated linear discriminant analyses (LDA). The leave-one-out procedure removes one specimen at a time, and predicts its classification using LDA functions computed on all the remaining specimens. The procedure is

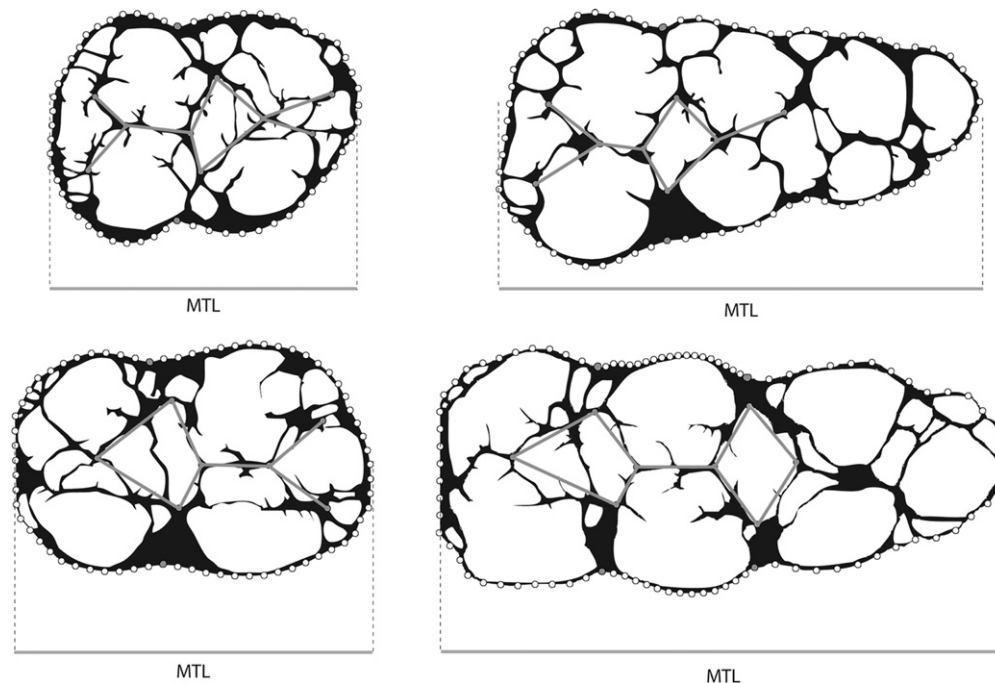


Fig. 1. Traditional measurement (MTL: Maximum Tooth Length) and location of the 2D landmarks (in grey), connected by lines to emphasize their relative positions, and sliding semi-landmarks (in white) along the outlines of the upper (above) and lower (below) second (left) and third (right) molars of *Sus scrofa*.

repeated for each specimen in the sample, each in turn being treated as an unknown, and avoids predicting a specimen on the basis of a function computed on data that includes the specimen itself – a type of circular reasoning that tends to spuriously inflate classification accuracy (see Kovarovic et al., 2011; and references therein). Classification accuracy is given by the percentage of specimens correctly assigned by the cross-validated LDA (cross-validated percentage, CVP).

3.2.2. Sensitivity to the number of shape predictors

If a LDA is used along with parametric tests of significance, the analysis cannot be performed unless sample size minus the number of groups is larger than the number of variables. If that is a minimum mathematical requirement for testing differences, it is also desirable that the number of variables is actually smaller than the sample size of the smallest group – a common rule-of-thumb for robust and reliable results (see references in Kovarovic et al., 2011). PCA can be used to reduce the number of predictors used in a LDA, by substituting the scores on the first axes of the PCA to the primary data. The number of PCs that have to be retained to achieve parsimony at the cost of a negligible loss of information is not fixed and a variety of methods has been proposed, some of which are reviewed in Zelditch et al. (2004), but none of them provides a definitive answer. Stepwise LDAs, and other methods that select the fewer number of predictors (PCs, in our case) maximizing the variability between groups, may appear convenient, as the number of variables for the classification is determined by the LDA itself. However, this approach has been strongly criticized because results are generally strongly sample-dependent and, therefore, unlikely to be generalized (Huberty and Hussein, 2003).

An alternative is to explore the sensitivity of results to the inclusion of varying numbers of PCs. This can be done by first selecting a sufficient number of PCs to accurately preserve the distances among the specimens in the shape space and then assessing whether including more or fewer variables appreciably changes classification accuracy (Viscosi and Cardini, 2011). Another possibility involves extensively exploring sensitivity by including just one or a few PCs and then progressively increasing their number (Baylac and Friess, 2005). This second approach has been used in the present study. LDAs were repeated using an increasing number of PCs (from 2 to 80 first PCs), which cumulatively explain more than 99% of total shape variance.

3.2.3. Sensitivity to unequal sample size

A common, but often neglected, issue of LDA is the potential influence of sample size heterogeneity on the results (Kovarovic et al., 2011; and references therein). Large differences in sample size across groups may lead to the largest sample dominating the pattern of variance covariance in the data (which is not an issue if the assumption of uniformity of variance – homoscedasticity is met), as well as to misinterpreting classification accuracy if results are not carefully examined. For instance, a 90% average cross-validated accuracy with two samples of 50 individuals each may not convey the same information as a 90% accuracy with two strongly unbalanced samples of 10 and 90 individuals. Indeed, in the first case random assignment is expected to yield a 50% CVP, but it will be higher in the second case. In this, as in other cases, the random chance threshold will therefore depend on the relative composition of unbalanced samples (Kovarovic et al., 2011). In general, uneven sample size renders the interpretation of results more difficult. In our dataset, the relative number of domestic pig and wild boar was highly uneven, with wild boars being three to five times more numerous than domestic pigs (Table 1). The sensitivity of the LDA to the number of specimens in each group can

also be assessed using randomization experiments. To this aim, we replicated our analyses with:

a) Perfectly **balanced groups** obtained by a random selection of a number of specimens in the largest sample (wild boars) equal to the sample size of the smallest group (domestic pig). If unequal sample size has little influence on cross-validation results, classification accuracy should be comparable in the random balanced groups to the accuracy estimated including all specimens with their initial unbalanced distributions. (b) **Samples with randomized group affiliation**. These groups are built by assigning specimens to one or the other group by chance. Thus, the observed pattern of group affiliation is intentionally removed before LDA and correct classification of specimens is only due to chance. By repeating this procedure multiple times, the average and range of percentages of individuals correctly classified by chance can be estimated – for a given sample size and set of predictors. This gives the empirical threshold above which classification accuracy is better than chance, which may be very different from 50%. The randomization procedure was performed on both the original dataset and on balanced group datasets. The randomized selections and re-sampling experiments were repeated 1000 times and the analysis outcomes summarized by the mean and upper and lower 95th percentiles.

3.2.4. Distribution of posterior identification probabilities

A specimen can be assigned to a group in a predictive LDA with higher or lower confidence, depending on its relative distance to the group means. The level of confidence in a LDA is estimated by the posterior probabilities of classification. The highest posterior probability determines the classification of the specimen, and in the case of two groups, this probability ranges from 50 to 100%. However, being classified in a group with, for instance, a 51% posterior probability does not suggest the same level of confidence as being classified in that group with a 99% probability. Indeed, different morphometric descriptors could be equally accurate in terms of the frequency of correctly classified specimens, but with different levels of confidence. The distributions of posterior probabilities were visualized with histograms based on MTL, CS and shape (20th first PCs). In these diagrams, a 90% posterior probability threshold was used to help target the proportion of individuals classified with the highest confidence. Predictive accuracy and confidence levels were also compared using Receiver Operator Characteristic Curves (ROC) and the corresponding areas under the curve (AUC) statistic. ROC are computed using posterior probabilities from the LDAs so that the highest curve, with the largest AUC, represents the best predictive model (Franklin et al., 2012; Sanfilippo et al., 2010).

Morphometric and statistical analyses were performed in R v 2.13.1 (R Development Core Team, 2011) using “Rmorph” (Baylac, 2012); the “ROCR” package (Sing et al., 2005) and newly designed functions (available under request).

4. Results

4.1. Cross validation percentages

4.1.1. Full samples

The two measures of size, MTL and CS, were highly correlated (correlations ranging between $r = 0.945$ and $r = 0.980$ depending on the dataset). The variation of MTL, CS and shape in wild and domestic pigs showed overlapping ranges (Fig. 2). LDAs correctly identified on average 77.9%–87.5% of specimens using MTL, 77.3%–85.6% using CS and 92.0%–93.0% using shape (for each tooth using the number of first PCs that maximize the CVP, ranging between 13 and 54) (Table 2, Fig. 3). There was a bias in the percentage of

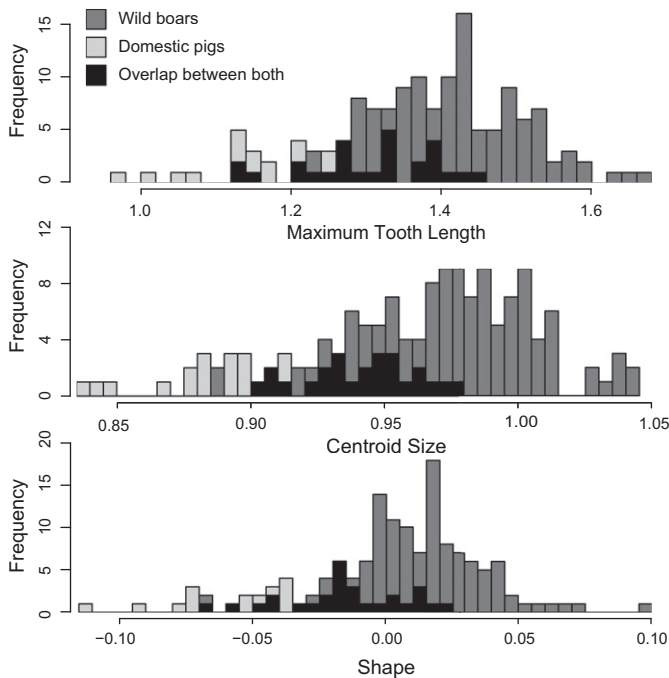


Fig. 2. Histograms showing frequency distributions of maximum tooth length, centroid size and shape for the lower M3. The shape variable corresponds to PC1 of the between group PCA.

misidentified specimens, with domestic pigs being misidentified up to three times as often as wild boar (Table 2). CVP was slightly higher using the second molars (M2) compared to the third (M3) (Table 2). LDAs of shape data were sensitive to the number of PCs included in the analyses, with CVP increasing with the number of PCs until ca. 10–15 PCs are included (Fig. 3).

4.1.2. Randomized unbalanced groups

Randomized-unbalanced CVP distributions based on sizes (MTL and CS) always fall below their respective observed CVP (Fig. 3). When the number of shape PCs included in the LDA was increased, CVP decreased (Fig. 3) from an average of 82%–74% for the upper M2, from 75% to 60% for the upper M3, from 83% to 75% for the lower M2 and from 75% to 61% for the lower M3. As expected for unbalanced sample sizes, CVP computed on randomized datasets were above 50% (Fig. 3). When more than ca. 5 PCs were included, the initial CVPs were always above those of randomized-unbalanced analyses.

4.1.3. Balanced groups

On average, balanced groups CVPs estimated using MTL ranged between 67% and 79% in average depending on the tooth, and varied between averages of 69% and 81% when CS was used

(Fig. 3). Using shape, on average, balanced groups CVPs increased rapidly with the number of PCs, until a plateau was reached when ca. 10–20 of the first PCs were analyzed (Fig. 3). Including more than 40–50 PCs led on average to a decrease in CVPs. This decrease was especially pronounced for third molars (Fig. 3b and c). Maximum of the means of the CVP distributions obtained for balanced sub-samples, paired with the maximum of the upper limits of the confidence intervals are show in Table 2.

4.1.4. Balanced-randomized groups

As expected, balanced-randomized CVP values (Fig. 3) varied around 50% average using all descriptors (MLT, CS and shape). Results obtained for balanced groups (4.1.3) were above CVPs from randomized balanced groups using size (MLT, CS). Using shape also CVPs from balanced groups tended to be larger than those from randomized balanced groups. The main exception was when more than the ca. 65 of the first M3 PCs were analyzed, which led to almost fully overlapping CVP distributions from balanced and balanced randomized groups (Fig. 3b and d). Using M2 shape (Fig. 3a and c), CVPs also tended to decrease when very many PCs were included, but this never led to overlapping distributions.

Comparing CVP from full samples and balanced ones revealed the sensitivity of CVP to unequal sample sizes. CVPs based on size measures were always above those from balanced samples. The classification bias (difference between the observed CVP and the 95th upper percentile of the balanced groups CVP distribution) was between 3% and 8% for MTL and between 1% and 5% for CS.

CVPs based on shape also were influenced by unequal sample size with CVPs from full samples generally larger than those from balanced samples. For upper M2 (Fig. 3a), whose full sample CVPs were always larger than those balanced data, the classification bias ranged from 2% to 15%, depending of the number of included PCs. For lower M2 (Fig. 3c), there was a small region of overlap in CVP ranges (ca between 13 and 25 PCs) and the positive bias ranged between 2% and 15%. For third molars (Fig. 3b and d), a positive classification bias was found only when more than the first 41 (lower M3) – 48 (upper M3) were analyzed and pronouncedly increased to become as large as ca. 23–28% when the 80 first PCs were included.

4.2. Distribution of posterior probabilities of identification

A high heterogeneity was observed in the distributions of posterior identification probabilities (Fig. 4). The distributions based on size measurements (MTL and CS), although not uniform, revealed no sharp tendency (Fig. 4a and b) and included only between 20% and 58% of specimens identified with probabilities above the 0.9 threshold for MTL, and between 23% and 65% for CS. In contrast, the distribution of posterior probabilities based on shape had a very pronounced asymmetry with a most evident peak in the highest probability range and between 76% and 87% of the

Table 2

Results of the LDA displaying the cross validation percentages (CVP) of correct classification for maximum tooth length (MTL), centroid size (CS) and shape as well as the percentages of miss-identified (missID) domestic pigs (DP) and wild boars (WB). For shape, the maximum CVP value is shown together with the corresponding number of N first PCs (NPCs) included. Results obtained for balanced sub-samples correspond to the maximum of means CVP values and the maximum of the upper limits of the confident intervals (CI).

	With full sample (unbalanced)						With balanced sub-samples						
	MTL			CS			Shape			MTL	CS	Shape	
	CVP	DPmissID	WBmissID	CVP	DPmissID	WBmissID	CVP	NPCs	DPmissID	WBmissID	Max-Mean CVP/Max-upper CI		
Upper M2	82.6	45	15	85.6	31	13	92.97	47	16	6	69.7/73.7	74.4/79.6	83.4/88.1
Upper M3	77.9	38	20	77.3	41	21	92.07	23	18	6	67.8/71.3	70.1/75	87.5/92.5
Lower M2	87.5	31	10	89.4	25	9	92.97	54	24	5	79.4/82	80.6/84	86.2/91.5
Lower M3	81.7	32	16	82.2	30	15	92.34	13	16	5	72.3/77.3	74.8/78	88.3/92.9

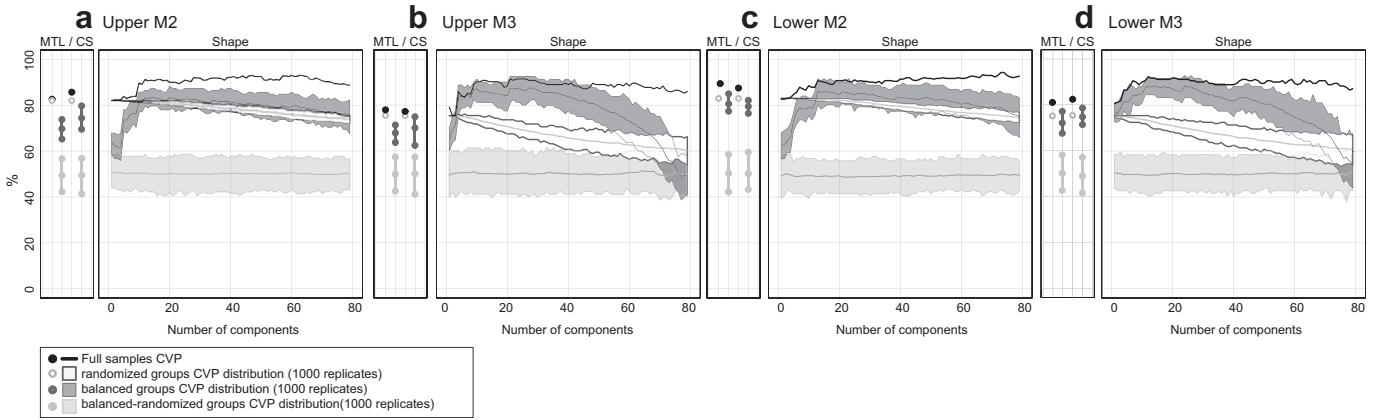


Fig. 3. Sensitivity of discriminant analyses to unequal sample size and to number of shape predictors. Classification accuracy (cross validation percentages (CVP)) of domestic and wild pigs based on maximum tooth length (MLT), centroid size (CS) and shape for upper M2 (a) and M3 (b), lower M2 (c) and M3 (d) using full samples data (black), groups of balanced effectives (dark grey), randomized groups of unbalanced effectives (white), and randomized groups of balanced effectives (light grey). Distributions of the CVP were visualized using means and 95th upper percentiles and from 1000 replicates of each analysis using subsamples and/or randomized data.

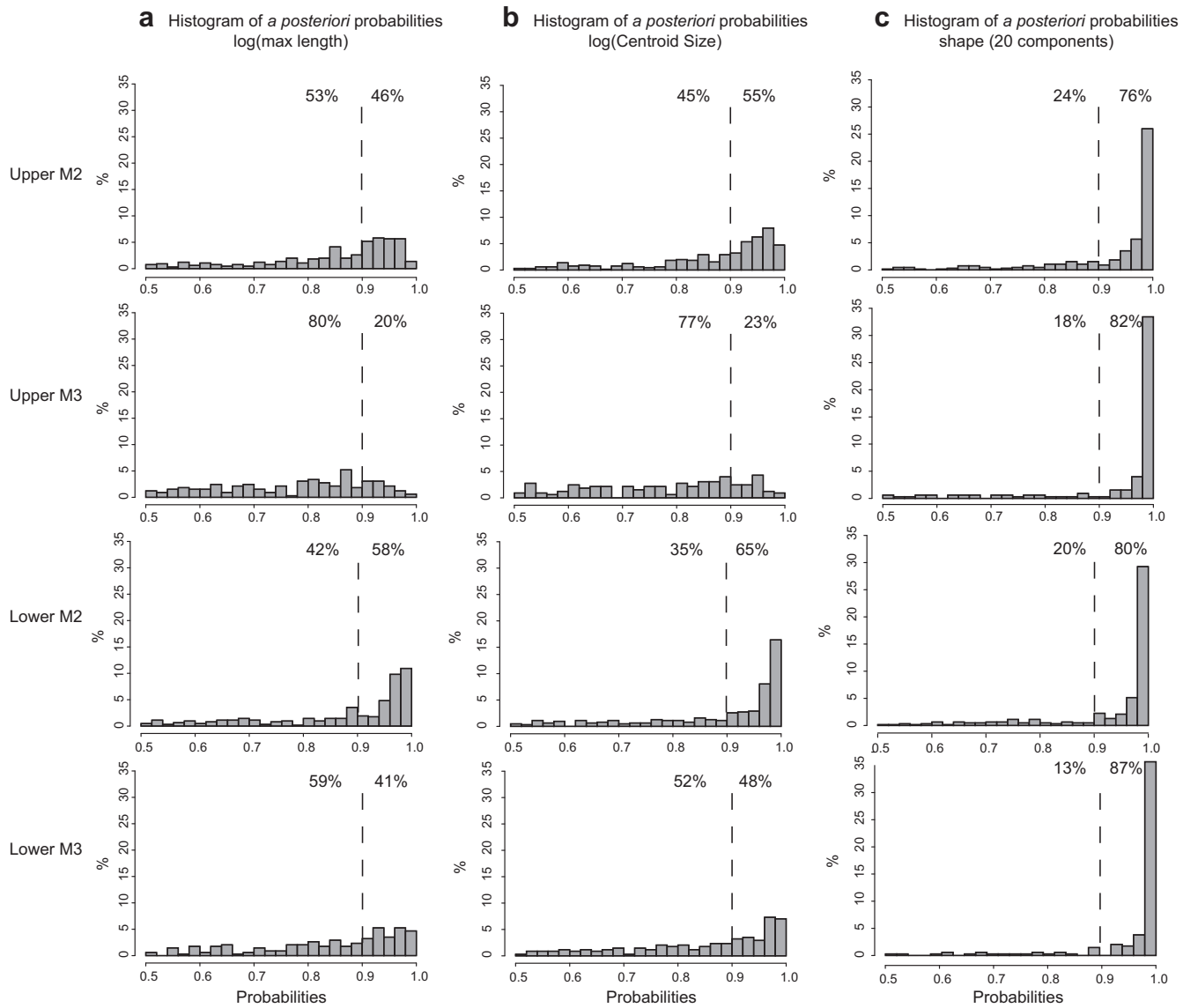


Fig. 4. Distributions of the a posteriori probabilities of identification based on (a) maximum tooth length (MTL), (b) centroid size (CS) and (c) shape for each tooth. The percentages of specimens with a posteriori probability of identification below and above the 0.9 threshold (dotted vertical line) are shown for each dataset, and can be seen as the percentages of unidentifiable and identifiable specimens respectively.

specimens identified with posterior probabilities higher than 0.9. Another way to summarize these results is by saying that 42%–80% of all the specimens could be identified with a fairly high degree of uncertainty using maximum length; 45%–77% using centroid size and only 13%–24% using shape (Fig. 4).

4.3. AUC

For each tooth, AUC analyses suggested the exact same ranking in discriminatory accuracy with MTL being the poorest descriptor in terms of confidence, followed by CS, and finally shape, where AUC was 1.1–1.3 times larger than that of size measures (Table 3, Fig. 5). This result was not dependent of the tooth analyzed since discriminatory accuracies for shape area was always above the discriminatory accuracy obtained for size descriptors (MTL and CS) (Fig. 5).

5. Discussion

5.1. Discriminant analysis: accuracy and bias

As expected, the number of predictors as well as group sample sizes can strongly affect the results of discriminant analyses. For instance, even when its purpose is purely predictive, more predictors do not always perform better. Indeed, increasing the number of variables will increase the accuracy of the discriminant analyses but only up to a point where adding more variables do not appreciably improve results. Actually, if adding too many variables further inflates classification accuracy in non-cross-validated LDAs (Kovarovic et al., 2011), an exceedingly large number of predictors compared to sample size may eventually reduce classification accuracy in cross-validated LDAs (Kovarovic et al., 2011; Jain and Waller, 1978).

The sensitivity analysis we performed indicates that using the first ~20 PCs of shape may be a good compromise between accuracy and parsimony. The number of PCs is clearly dataset dependent, may vary with the type of predictor, number of landmarks, sample composition and the magnitude of differences. Progressively increasing the number of PCs until the percentage of correct cross-validation reaches a plateau (see also Baylac and Friess, 2005) may offer an interesting option to objectively decide the minimum number of predictors that should be included in LDA to achieve maximum discrimination.

As expected, in all analyses, unbalanced samples bias the cross-validation percentages in favour of the larger group. Our initial datasets had 3–5 times more wild boar than domestic pigs, which produce chance CVP distribution approximately between ca. 60 and 80% on average, i.e. well above the 50% expected for two balanced samples. When analyses from full samples and balanced groups are compared, a CVP positive bias was consistently found for size descriptors. In contrast, results from shape data were less strongly affected by unbalanced sample sizes with the main exception of the analyses using very many PCs. Indeed, the bias in CVP classification generally

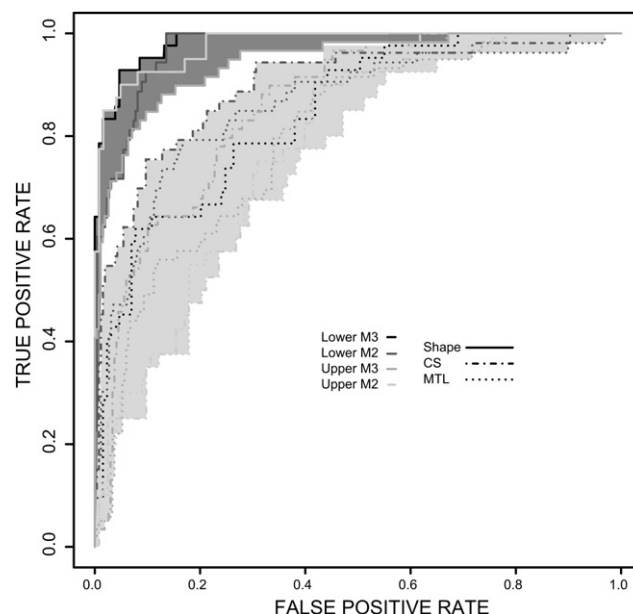


Fig. 5. Comparison of the ROC curves obtained for the four teeth, for shape (distribution in dark grey) and size descriptors (MTL and CS, distribution in light grey).

increased with the number of PCs. This seems evident for second molars, especially upper ones, across the whole range of PCs. However, when more than 50–60 PCs are included, the bias becomes much stronger in third molars.

The discrepancy between our analyses on second and third molars could, however, be largely due to the number of specimens and/or the number of landmarks analyzed. Indeed, in this study, third molars were less numerous than second molars which are only found in older specimens. Also, because third molars have more complex shapes, they require a larger number of landmarks and semi-landmarks for an accurate description, which inevitably generates more variables. The observation that classification accuracy (i.e., CVP) deteriorates when a very large number of PCs is included especially in third molars, leading to a larger discrepancy between CVPs of full and balanced samples (bias), is consistent with the smaller sample size used in the balanced analyses of third molars. As the maximum number of PCs cannot be more than total sample size minus one, using balanced samples of 40–42 specimens per group in respectively upper and lower M3 analyses, the largest number of PCs cannot be more than 79–83 whereas in M2 balanced analyses PCs are more than 100. This suggests that, despite the larger number of points used to capture the complex shape of M3, PCs from 50 to 80 pickup random variation (e.g., measurement error) in M3 (adding ‘noise’ to the DA and reducing CVP) whereas the same PCs are still informative in M2, as the random variation is likely accounted for by PCs higher than the 80th, which is the highest order PC in our analysis. Indeed, if balanced sample analyses (not shown) are done using the same sample size for all teeth (i.e., 40, which is the size of the smallest sample among all datasets), the trend in M2 CVPs shows the same pattern as in M3, with a sharp reduction when the last ca. 20 PCs are added.

Overall, this series of analyses indicates that performing first a PCA on predictors, in our case the shape coordinates, and then selecting an appropriate number of PCs in the range corresponding to the CVP plateau, may help not only to maximize the cross-validated accuracy but also to reduce the ‘noise’ in the data, as well as the positive bias in classification accuracy when groups have different sample sizes. However, even when the number of

Table 3

Area under the curve (AUC) of the receiver operating characteristic (ROC) statistic computed from posterior identification probabilities for each tooth and calculated for maximum tooth length (MTL), centroid size (CS) and shape. In bold the higher AUC value for each tooth.

	Upper M2	Upper M3	Lower M2	Lower M3
MTL	0.798	0.755	0.859	0.843
CS	0.857	0.777	0.898	0.870
Shape	0.945	0.978	0.971	0.985

predictors is carefully selected in relation to sample size and accuracy, a positive bias in CVPs is most likely to be found, as in our case study, and a random chance baseline should be estimated to interpret how good a classification is (Kovarovic et al., 2011) and, if possible, the sensitivity of results to uneven sampling explored using balanced subsamples.

5.2. Identifying wild and domestic pigs from molar size and shape

Modern wild and domestic pigs from the West Palaearctic largely overlap in their molar size and shape. In accordance with previous studies (see also Jarman, 1976; Payne and Bull, 1988), size and shape variation in the two groups is not dichotomous, which implies inevitable inaccuracies in classification. This makes it even more important to minimize errors during the identifications and to find the most accurate descriptors to discriminate domestic and wild pigs. This is especially crucial when tracking pig domestication in the archaeological record with the purpose of inferring when and where pigs were domesticated. To this aim, one has to bear in mind that the implicit assumption that modern samples can provide a good model for ancient populations may not necessarily always hold true. Thus, error rates using discriminant functions derived from contemporary material may be higher when applied to archaeological material. Our study provides an example of how to build an accurate numerical classification with valid probabilistic assessment of groups. The accuracy of the classification when applied to archaeological material could be tested on specimens for whom ancestry has been established using independent evidence, such as genetic markers, isotopes, kill-off patterns and the archaeological contexts. Results can be used to refine and improve the classification functions.

The classification between wild and domestic forms in our analysis had an average cross-validated accuracy of about 85% (range: 77.3–91.2%), results which were significantly better than random chance accuracy as indicated by CVPs above the chance estimated accuracy in almost all DAs.

Size analyses appeared to be more sensitive to unbalanced sample size than shape. Indeed, CVP obtained for unbalanced samples were always higher than for balanced samples when sizes descriptors are considered. In contrast, using shape, unbalanced samples did not necessarily largely inflate CVP especially when an appropriate (i.e., neither too few nor too many, as discussed before) number of PCs was included in the analysis. More importantly, a discriminant function should not only correctly classify most of the specimens, but it should also do so with high posterior for the classification to be both accurate and reliable. Posterior probabilities for all specimens should be shown. This is often impractical. However, one can easily explore the distribution of posterior probabilities in a sample to provide information on the degree of confidence one can have on average in the group predictions. That this is highly informative is evident from our study; employing a 0.9 posterior probability threshold would have led to a large number of unidentified specimens using size measures. This finding is consistent with previous work (Franklin et al., 2012) suggesting that shape data are more reliable in group discrimination compared to simple size measures. Size can be recorded more rapidly than shape, but does not provide a high degree of confidence even when its predictive accuracy is apparently high. Conversely shape provides more convincing results and at the same time is easily measured on low cost digital images when objects are well approximated in two dimensions.

Due to clear sexual dimorphism in molar size in wild boars (Evin et al., in prep.) – with males being larger than females – a classification based solely on size will likely assign males to wild boar and females to domestic pigs independently of whether

they really are wild or domestic. Moreover, since the sex of the specimen is rarely identified in archaeological material, separate analysis to identify male and female is, therefore, impossible. This bias, in contrast, is unlikely to occur using molar shape, because shape does not reflect sexual dimorphism (Cucchi et al., 2011b; Evin et al. in prep.).

Geography is another factor of phenotypic variability within wild boar (e.g. Groves, 1981; Albarella et al., 2009). However, a further study (not shown), looking at differences between wild and domestic form – and using a single wild population – provided similar results, suggesting that geographic factor do not seem to interfere with the overall differentiation between wild and domestic pigs. It would, however, be interesting to specifically analyze geographical variation within wild boar in the context of domestication, as well as to extend this study to the East Palaearctic where another centre of pig domestication has been shown (Larson et al., 2005).

Our study, therefore, consistently indicates that molar shape, captured using geometric morphometrics techniques, provides an accurate method to identify the domestic status of archaeological specimens and is superior to traditional size measures. That different teeth performed equally well also suggests that the choice of structure to be analyzed may be less important than the selection of an accurate descriptor such as shape compared to less reliable and potentially biased size data.

6. Conclusion

In conclusion, our study shows that size is a poor indicator of wild or domestic status, even when the refined comprehensive quantification of centroid size is used, and results from comparisons of size measures should be interpreted with the greatest caution. Shape is not only accurate in terms of predicting whether teeth derived from wild or domestic pigs but also provides a high degree of confidence in those predictions. This indicates that shape analysis is a powerful tool in zooarcheological research and suggests that studies similar to the one we have performed on pig teeth should be carried on in other taxa such as cattle and auroch, or sheep and goat, for which taxonomical identification is often challenging using traditional techniques, and extended to other skeletal tissues such as epiphyses of long bones, or other traditionally diagnostic parts. Furthermore, while predictive linear discriminant analysis is useful, well established and a relatively simple method to use for group discrimination, its results must be both cross-validated and carefully examined with regard to confidence (i.e. distribution of posterior probabilities), effect of heterogeneous sample size and number of predictors.

Acknowledgements

We thank the many institutions and individuals that provided sample material and access to collections, especially the curators of the Museum für Naturkunde, Berlin; Muséum National d'Histoire Naturelle, Paris; Muséum d'Histoire Naturelle, Genève; Museum für Haustierkunde, Halle; National Museum of Natural History, Washington; The Field Museum, Chicago and The American Museum of Natural History, New-York. We also thank Paul Sanfilippo (University of Melbourne) for his help with the R implementation of the ROC/AUC analysis and Vincent Debat, Nelly Gidaszewski and Mathieu Joron (CNRS-MNHN, Paris) for early comments on the manuscript and Ardern Hulme-Beaman for the teeth drawings. We are grateful to Krish Seetah and an anonymous referee for their comments on this manuscript. This work was supported by NERC (grant number NE/F003382/1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jas.2012.08.005>.

References

- Adams, D., Rohlf, F.J., Slice, D., 2004. Geometric morphometrics: ten years of progress following the “revolution”. *Ital. J. Zool.* 71, 5–16.
- Albarella, U., Davis, S.J.M., Detry, C., Rowley-Conwy, P., 2005. Pigs of the “far west”: the biometry of *Sus* from archaeological sites in Portugal. *Anthropozoologica* 40, 27–54.
- Albarella, U., Dobney, K., Rowley-Conwy, P., 2006. The domestication of the pig (*Sus scrofa*): new challenges and approaches. In: Zeder, M.A., Decker-Walters, D., Bradley, D., Smith, B.D. (Eds.), *Documenting Domestication: New Genetic and Archaeological Paradigms*. University of California Press, Berkeley, pp. 209–227.
- Albarella, U., Dobney, K., Rowley-Conwy, P., 2009. Size and shape of the Eurasian wild boar (*Sus scrofa*), with a view to the reconstruction of its Holocene history. *Environ. Archaeol.* 14, 103–136.
- Albarella, U., Payne, S., 2005. Neolithic pigs from Durrington Walls, Wiltshire, England: a biometrical database. *J. Archaeol. Sci.* 32, 589–599.
- Baylac, M., 2012. Rmorph: a R Geometric and Multivariate Morphometrics Library.
- Baylac, M., Friess, M., 2005. Fourier descriptors, procrustes superimposition and data dimensionality: an example of cranial shape analysis in modern human populations. In: Slice, D.E. (Ed.), *Modern Morphometrics in Physical Anthropology*. Springer-Verlag, New York, pp. 145–165.
- Bignon, O., Baylac, M., Vigne, J.-D., Eisenmann, V., 2005. Geometric morphometrics and the population diversity of Late Glacial horses in Western Europe (*Equus caballus arcelini*): phylogeographic and anthropological implications. *J. Archaeol. Sci.* 32, 375–391.
- Boessneck, J., von den Driesch, A., 1978. The significance of measuring animal bones from archaeological sites. In: Meadow, R.H., Zeder, M.A. (Eds.), *Approaches to Faunal Analysis in the Middle East*. Bulletin N°2. Peabody Museum, Harvard University, pp. 25–40.
- Boessneck, J., Jéquier, J.P., Stampfli, H.R., 1963. Seeberg, Burgäschisee-Süd. Band 3: Die Tierreste (Acta Bernensia II/3).
- Bookstein, F.L., 1991. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.
- Bookstein, F.L., 1996. Biometrics, biomathematics and the morphometric synthesis. *Bull. Math. Biol.* 58, 313–365.
- Bookstein, F.L., 1997. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Image Anal.* 1, 225–243.
- Bökönyi, S., 1974. *History of Domestic Mammals in Central and Eastern Europe*. Akadémiai Kiado, Budapest.
- Bull, G., Payne, S., 1982. Tooth eruption and epiphysal fusion in pigs and wild boar. In: Wilson, B., Grigson, C., Payne, S. (Eds.), *Ageing and Sexing Animal Bones from Archaeological Sites*. BAR British Series vol. 109, pp. 55–71.
- Cucchi, T., 2008. Uluburun shipwreck stowaway house mouse: molar shape analysis and indirect clues about the vessel's last journey. *J. Archaeol. Sci.* 35, 2953–2959.
- Cucchi, T., Fujita, M., Dobney, K., 2009. New insights into pig taxonomy, domestication and human dispersal in Island South East Asia: molar shape analysis of *Sus* remains from Niah caves, Sarawak. *Int. J. Osteoarchaeol.* 19, 508–530.
- Cucchi, T., Balasescu, A., Bem, C., Radu, V., Vigne, J.-D., Tresset, A., 2011a. New insights into the invasive process of the eastern house mouse (*Mus musculus musculus*): evidence from the burnt houses of Chalcolithic Romania. *The Holocene* 21, 1195–1202.
- Cucchi, T., Hulme-Beaman, A., Yuan, J., Dobney, K., 2011b. Early Neolithic pig domestication at Jiahui, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. *J. Archaeol. Sci.* 38, 11–22.
- Davis, S.J.M., 1981. The effects of temperature change and domestication on the body size of late Pleistocene to holocene mammals of Israel. *Paleobiology* 7, 101–114.
- Dobney, K., Larson, G., 2006. Genetics and animal domestication: new windows on an elusive process. *J. Zool.* 269, 261–271.
- Ervynck, A., Dobney, K., Hongo, H., Meadow, R., 2001. Born free? New evidence for the status of *Sus scrofa* at neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey). *Paléorient* 27, 47–73.
- Flannery, K.V., 1982. Early domestication in the fertile crescent a retrospective look. In: Young Jr., T.C., Smith, P.E.L., Mortensen, P. (Eds.), *The Hilly Flanks: Essays on the Prehistory of Southwestern Asia*. Studies in Ancient Oriental Civilization, vol. 36. Orient Inst. Univ. Chicago, Chicago, pp. 163–188.
- Franklin, D., Cardini, A., Flavel, A., Kuliukas, A., 2012. The application of traditional and geometric morphometric analyses for forensic quantification of sexual dimorphism: preliminary investigations in a Western Australian population. *Int. J. Legal. Med.* <http://dx.doi.org/10.1007/s00414-012-0684-8>
- Goodall, C.R., 1995. Procrustes methods in the statistical analysis of shape revisited. In: Mardia, K.V., Gill, C.A. (Eds.), *Current Issues in Statistical Shape Analysis*. University of Leeds Press, Leeds, pp. 18–33.
- Groves, C., 1981. *Ancestors for the Pigs: Taxonomy and Phylogeny of the Genus Sus* (Technical Bulletin No. 3, Department of Prehistory, Research School of Pacific Studies). Australian National University, Canberra.
- Gunz, P., Mitteroecker, P., Bookstein, F.L., 2005. Semi-landmarks in three dimensions. In: Slice, D. (Ed.), *Modern Morphometrics in Physical Anthropology*. Kluwer Academic/Plenum Publishers, New York, pp. 73–98.
- Hongo, H., Pearson, J., Öksüz, B., Ilgezdi, G., 2009. The process of ungulate domestication at Çayönü, Southeastern Turkey: a multidisciplinary approach focusing on *Bos sp.* and *Cervus elaphus*. *Anthropozoologica* 44, 63–78.
- Huberty, C.J., Hussein, M.H., 2003. Some problems in reporting use of discriminant analyses. *J. Exp. Educ.* 71, 177–191.
- Jain, A.K., Waller, W.G., 1978. On the optimal number of features in the classification of multivariate Gaussian data. *Pattern Recognit.* 10, 365–374.
- Jarman, M.R., 1976. Prehistoric economic development in sub-Alpine Italy. In: Sieveking, G.D.G., Longworth, I.H., Wilson, K.E. (Eds.), *Problems in Economic and Social Archaeology*. Duckworth, London, pp. 375–399.
- Kovarovic, K., Aiello, L.C., Cardini, A., Lockwood, C.A., 2011. Discriminant function analyses in archaeology: Are classification rates too good to be true? *J. Archaeol. Sci.* 38, 3006–3018.
- Larson, G., Cucchi, T., Fujita, M., Masakatsu, M., Matisoo-Smith, E., Robins, J., Anderson, A., Rolett, B., Spriggs, M., Dolman, G., Kim, T.-H., Thuy, N.T.D., Randi, E., Doherty, M., Due, R.A., Bollt, R., Djubiantono, T., Griffin, B., Intoh, M., Keane, E., Kirch, P., Li, K.-T., Morwood, M., Pedriña, L.M., Piper, P.J., Rabett, R.J., Shooter, P., Van den Bergh, G., West, E., Wickler, S., Yuan, J., Cooper, A., Dobney, K., 2007. Phylogeny and ancient DNA of *Sus* provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proc. Natl. Acad. Sci. U S A* 104, 4834–4839.
- Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L., Cooper, A., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307, 1618–1621.
- Mayer, J.J., Novak, J.M., Brisbin, I., 1998. Evaluation of molar size as a basis for distinguishing wild boar from domestic swine: employing the present to decipher the past. In: Nelson, S.M. (Ed.), *Ancestors for the Pigs: Pigs in Prehistory*. MASCA, Philadelphia, pp. 39–53.
- Meadow, R.H., 1989. Osteological evidence for the process of animal domestication. In: Clutton-Brock, J. (Ed.), *The Walking Larder: Patterns of Domestication, Pastoralism, and Predation*. Unwin and Hyman, London, pp. 80–90.
- O'Regan, H.J., Kitchener, A.C., 2005. The effects of captivity on the morphology of captive, domesticated and feral mammals. *Mammal Rev.* 35, 215–230.
- Payne, S., Bull, G., 1988. Components of variation in measurements of pig bones and teeth, and the use of measurements to distinguish wild from domestic pig remains. *Archaeozoologia* II, 27–66.
- Perez, S.I., Bernal, V., Gonzalez, P.N., 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *J. Anat.* 208, 769–784.
- Peters, J., Helmer, D., von den Driesch, V., Segui, M.S., 1999. Early animal husbandry in the Northern Levant. *Paléorient* 25, 27–48.
- Price, E.O., 2002. *Animal Domestication and Behavior*. CABI, Wallingford, UK.
- R Development Core Team, 2011. *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rohlf, F.J., 2010a. tpsDig. In: *Ecology and Evolution*. State University of New York at Stony Brook, Stony Brook, NY.
- Rohlf, F.J., 2010b. *Relative Warps*. In: *Ecology and Evolution*. State University of New York at Stony Brook, Stony Brook, NY.
- Rohlf, F.J., Slice, D., 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. *Syst. Biol.* 39, 40–59.
- Rohlf, J.F., Marcus, L.F., 1993. A revolution in morphometrics. *Trends Ecol. Evol.* 8, 129–132.
- Rowley-Conwy, P., Albarella, U., Dobney, K., 2012. Distinguishing wild boar from domestic pigs in prehistory: a review of approaches and recent results. *J. World Prehist.* 25, 1–44.
- Rüttimeyer, L., 1862. *Die fauna der pfahlbauten der schweiz*. Birkli, Basel.
- Sampson, P.D., Bookstein, F.L., Sheenan, F.H., Bolson, E.L., 1996. Eigenshape analysis of left ventricular outlines from contrast ventriculograms. In: Marcus, L.F., Corti, M., Loy, A., Naylor, G.J.P., Slice, D.E. (Eds.), *Advances in Morphometrics*. Plenum, New York and London, pp. 211–233.
- Sanfilippo, P.G., Cardini, A., Sigal, I.A., Ruddle, J.B., Chua, B.E., Hewitt, A.W., Mackey, D.A., 2010. A geometric morphometric assessment of the optic cup in glaucoma. *Exp. Eye Res.* 91, 405–414.
- Seetah, T.K., Cardini, A., Miracle, P.T., 2012. Can morphospace shed light on cave bear spatial-temporal variation? Population dynamics of *Ursus spelaeus* from Romulodova pećina and Vindija, (Croatia). *J. Archaeol. Sci.* 39, 500–510.
- Sheets, H., David, Covino, K.M., Panasiewicz, J.M., Morris, S.R., 2006. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. *Front. Zool.* 3, 15.
- Sheets, H.D., Keonho, K., Mitchell, C.E., 2004. A combined landmark and outline-based approach to ontogenetic shape change in the Ordovician Trilobite *Triarthrus becki*. In: Elewa, A.M.T. (Ed.), *Morphometrics: Applications in Biology and Paleontology*. Springer, New York, pp. 67–81.
- Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941.
- Valenzuela-Lamas, S., Baylac, M., Cucchi, T., Vigne, J.-D., 2011. House mouse dispersal in Iron Age Spain: a geometric morphometrics appraisal. *Biol. J. Linn. Soc. Lond.* 102, 483–497.
- Vigne, J.-D., Peters, J., Helmer, D., 2005. *The First Steps of Animal Domestication: New Archaeozoological Techniques*. Oxbow Books, Oxford.
- Viscosi, V., Cardini, A., 2011. Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners. *PLoS ONE* 6, e25630.
- Zeder, M.A., Bradley, D.G., Emshwiller, E., Smith, B.D., 2006. *Documenting Domestication, New Genetic and Archaeological Paradigms*. University of California Press, Berkeley.
- Zelditch, M., Swiderski, D.L., Sheets, H.D., Fink, W.L., 2004. *Geometric Morphometrics for Biologists: a Primer*. Elsevier Academic Press, New York and London.