Using traditional biometrical data to distinguish West Palearctic wild boar and domestic pigs in the archaeological record: New methods and standards

Allowen EVIN^{a, b}, Thomas CUCCHI^{b, a}, Gilles ESCARGUEL^c, Joseph OWEN^a, Greger LARSON^d, Una STRAND VIDARSDOTTIR^e, Keith DOBNEY^a

^aDepartment of Archaeology, University of Aberdeen, St.Mary's Building, Elphinstone Road, United Kingdom. a.evin@abdn.ac.uk; jowen@abdn.ac.uk; keith.dobney@abdn.ac.uk.

^bCNRS-Muséum National d'Histoire Naturelle, UMR 7209, Archéoozoologie, histoire des sociétés humaines et des peuplements animaux, 55 rue Buffon, 75005 Paris, France. evin@mnhn.fr; cucchi@mnhn.fr.

- ^c Laboratoire de Géologie de Lyon, UMR-CNRS 5276, Université Claude Bernard Lyon 1, Bât. Géode, Bvd. du 11 Novembre 1918, 69622, Villeurbanne Cedex, France. gilles.Escarguel@univ-lyon1.fr.
- ^d Durham Evolution and Ancient DNA, Department of Archaeology, University of Durham, South Road, Durham DH1 3LE, United Kingdom. greger.larson@durham.ac.uk.
- ^eDepartment of Anthropology, University of Durham, South Road, Durham DH1 3LE, United Kingdom. una.vidarsdottir@durham.ac.uk.

Keywords: cut-off, error risk, archaeology, domestication, Sus scrofa, molar size

Abstract

Traditionally, the separation of domestic pig remains from those of wild boar in zooarchaeological assemblages has been based on the comparison of simple size measurements with those from limited numbers of modern or archaeological reference specimens and then applying poorly defined cut-off values to make the identification calls. This study provides a new statistical framework for the identification of both domestic and wild *Sus scrofa* using standard molar tooth lengths and widths from a large modern comparative collection consisting of 407 West Palearctic wild boar and domestic pigs. Our study continues to rely upon so-called 'cut-off' values that correspond to the optimal separation between the two groups, but based upon a measure and visualisation of the error risk curves for erroneous identifications. On average, wild boar have larger teeth than domestic pigs and cut-off values were established for maximum tooth length and width,

respectively as follows: 2.39 cm and 1.85 cm for second upper molar, 3.69 cm and 2.13 cm for third upper molar, 2.26 cm and 1.50 cm for second lower molar, 3.79 cm and 1.75 cm for third lower molar. Specimens below and above these cut-offs are most likely to be, respectively, domestic pig and wild boar and the risk of providing a wrong identification will depend on the distance to the cut-off value following a relative risk curve. Although likely containing high risk of inherent statistical error, nonetheless this basic metrical identification-tool (based only on recent specimens), is here shown to correctly re-identify 94% of the Neolithic pigs from Durrington Walls (England) as domestic pig. This tool could be employed not only to systematically re-evaluate previous identifications of wild or domestic *Sus scrofa*, but also to establish new identifications where more powerful and reliable approaches such as Geometric Morphometrics cannot be applied.

Introduction

The domestication of certain plants and animals at the beginning of the Holocene epoch beginning some 10,000 years ago heralded perhaps one of the most significant biocultural steps in the history of mankind. As a result, the study of the origins and spread of farming, through the palaeobotanical and zooarchaeological record provides the baseline datasets for understanding not only crucial aspects of complex evolutionary history of the species involved in their transition from wild to domesticated organisms, but also crucial biocultural evidence linked with the shift from hunting and gathering to early farming.

Separating 'wild' from 'domestic' in the early zooarchaeological record is therefore one of the most important challenges facing researchers studying domestication, yet it remains one of the most difficult. Charles Darwin was the first to notice a range of morphological and phenotypic traits common to many domestic animals yet different to their wild ancestors (Darwin 1868), These include e.g. an obvious decrease in brain and body size, changes in some body proportions, and modification of external morphological characters such as emergence of piebald coat colour, wavy or curly hair, rolled and shortened tails, or floppy ears (Trut 1999, O'Regan and Kitchener 2005). Many of the phenotypic and behavioural changes linked with domestication are inaccessible from zooarchaeological assemblages, where only skeletal and dental remains are available for study. New techniques of ancient DNA analyses are providing novel information about phenotype (e.g., the coat colour of mammals, Ludwig et al. 2009), but these data are not routinely available, due to poor preservation and analytical costs. The zooarchaeological record is often very fragmented, and usually dominated by teeth that

are more easily identified using morphological or biometric criteria (von den Driesch 1976, Payne and Bull 1988). Identifying domestication using distinct morphological markers is therefore of prime interest for zooarchaeologists and is one of the principal approaches used extensively to do so over the last decades.

In the west Palaearctic, domestic forms of three taxa are particularly difficult to recognise in the archaeological record: cows (*Bos taurus*), dogs (*Canis familiaris*) and pigs (*Sus scrofa*). These three species are more difficult to recognise than, e.g., sheep (*Ovis aries*) or goat (*Capra aegagrus hircus*), because of the ubiquitous presence of their wild ancestors across western Eurasia (Aulagnier et al. 2008). For instance, we now know from recent ancient DNA research that the history of pig domestication is complex, and includes several processes of both local domestication, dispersal and introgression of wild and domestic forms (e.g., Larson et al. 2005, 2007, Ottoni et al. 2013, Larson & Burger 2013, Krause-Kyora et al. accepted). Objective and accurate criteria are therefore necessary to disentangle the wild and domestic forms of these species during the Holocene. In this context, the identification of wild and domestic pigs from archaeological remains have been commonly assessed using traditional size measurments of teeth and bones (e.g. Vigne et al. 2005). For pigs (and other domestic taxa), small individuals are commonly identified as 'domestic' and large as 'wild' (Albarella et al. 2006, Rowley-Conwy et al. 2012) even if an important overlap in size does exist between the two groups (e. g. Payne & Bull 1988, Evin et al. 2013).

Identification of zooarchaeological remains is often undertaken using a framework of 'reference' individuals of known geographic origin and or wild/domestic status. To identify the biometrical affinity of *Sus scrofa* remains from archaeological sites in Europe, the most commonly used reference datasets are either modern Turkish wild boar (Payne and Bull 1988) or late Neolithic domestic pigs from the UK site of Durrington Walls (Albarella and Payne 2005). These biometrical datasets are first and foremost limited both in their geographic and temporal extent and so their relevance or applicability to zooarchaeological collections from differeing times or places should be questioned. Additionally, a wild boar reference dataset should consist of more than a single population since wild boar are known to be very variable in size across their geographic range (e.g., Groves, 1981; Albarella et al., 2009; Rowley-Conwy et al 2012).

More recently, studies have employed the more powerfull approach of geometric morphometrics to study morphological change in pig domestication (e.g., Cucchi et al. 2009, 2011, Evin et al. 2013). In one study, molar size was shown to be a much poorer indicator of wild or domestic status in modern Sus scrofa than shape variables (Evin et al. 2013). Indeed, the size of wild and domestic modern West Palaeactic pigs largely overlaps and does not show a bimodal distribution, which implies inevitable high classification error rates (Payne & Bull 1988, Evin et al. 2013). On the other hand, geometric morphometric analyses of molar shape provide much better identification paired with higher classification probabilities. Sadly, geometric morphometric approaches have yet to become routinely applied in zooarchaeological studies. When compared to traditional techniques, they require learning new techniques about multivariate statistics and morphometrics, usually more sophisticated and expensive tools for data acquisition, and they require more time to measure and analyse the collections than traditional methodologies used by zooarchaeologists over the last decades of research. In addition, geometric morphometric (GMM) techniques do not allow the reexamination of previously published data without full re-analysis of the original archaeological (and relevant reference) specimens.

From this perspective, this study aims to provide:

- a new biometric framework for size measurments of modern domestic pig breeds and wild boars from a large geographic area, in order to provide descriptive statistics based on larger datasets than those already available;
- 2) statistically-controlled and more objective criteria to identify wild and domestic pigs using standard measurements of Maximum Tooth Length (MTL) and Maximum Tooth Width (MTW) on the 2^{nd} and 3^{rd} upper and lower molars.

This approach relies on the definition of cut-off values that correspond to the optimal separation between the two groups based on a measure and visualisation of the error risk curves for erroneous identifications.

In order to validate the identification-tool proposed, the results obtained were compared to the published measurements of the *Sus* specimens from the Late Neolithic site of Durrington Walls (Wiltshire, southern England), for which the measurements were published with the aim of being used as a standard of archaeological domestic pigs (Albarella & Payne 2005).

Material

The comparative specimens used in this study are the same as those in Evin et al. (2013), and correspond to 407 modern wild and domestic specimens represented by 327 upper M2 (M²), 163 upper M3 (M³), 311 lower M2 (M₂) and 171 lower M3 (M₃) (table 1). Wild boar specimens originate from North Africa (Algeria, Morocco), Europe (France, Switzerland, Germany, Poland), Near East (Turkey, Syria, Iran, Iraq) and Russia (see SI-1 for sample sizes). Domestic specimens belong to the following breeds: Berkshire, Cornwall, Deutsches Edelschwein, Corsican, Sardinian, Tamworth, Middle White, Hannover Braunschweiger Landschwein, Veredeltes Landschwein and Mangalitza (see SI-2 for sample sizes). All specimens are adults and from both sexes. Standard zooarchaeological tooth measurements – i.e. Maximum Tooth Length (MTL) and Maximum Tooth Width (MTW) - measured in centimetres, were extracted from the geometric morphometric data presented in Evin et al. (2013). MTL and MTW were measured as the distance, automatically extracted, between the Cartesian coordinates of the most anterior and the most posterior semi-landmarks, and the most labial and lingual semi-landmarks, respectively. To confirm that the Estimated MTL (EMTL) and Estimated MTW (EMTW) are accurate estimates of the traditional measurements of the MTL and MTW, direct and estimated measures of lengths and widths (MTL-EMTL and MTW-EMTW) were compared for a subsample of 100 specimens based on pictures using TpsDig2 v2.16 (Rohlf 2010).

In their paper on the Neolithic pigs from Durrington Walls, Albarella and Payne (2005) published not only the summary of the measurements but also the full dataset, allowing direct comparisons with our results. This dataset contains 82 MTL and 79 MTW of M^2 , 39 MTL and 45 MTW of M^3 , 81 MTL and 84 MTW of M_2 and 39 MTL and 42 MTW of M_3 . When two width measurements were available for one tooth the largest was used so as to be consistent with our own measurements.

Methods

Comparison of the estimated and the traditional variables

Linear least-square regressions were computed between MTL and EMTL and between MTW and EMTW, with MTL and MTW used as explanatory variables, and EMTL and EMTW as the response variables. To assess whether EMTL and EMTW are unbiased estimates of MTL and MTW, respectively, we calculated the 95% confidence intervals of the respective slope and intercept obtained for each regression, a perfect estimation corresponding to a slope of 1

and an intercept of 0. The relationships between MTL/EMTL and MTW/EMTW were then visualized using bivariate graphics.

Differences between wild boar and domestic pigs

Differences in MTL and MTW between modern wild and domestic pigs for each cheek tooth were tested using the nonparametric Wilcoxon's test and visualized with boxplots. A boxplot graphically represents the median and the four quartiles that contain each 25% of the values. Confidence intervals of the medians were also visualized in the boxplot by notches around the median. A non-overlap between notches of two plots is strong evidence that the two medians differ (Chambers et al. 1983).

Cut-off values and error risk for identifying wild & domestic Sus

The cut-off values separating modern wild and domestic pigs were estimated for each measurement and tooth following the protocol of Favre et al. (2008) and using the OpCut-Location v. 1.0 IDL® program developed by one of the authors (G. E.; Favre et al. 2008). The cut-off value was calculated from the means and standard deviations of the two a priori defined sets (wild boar and domestic pigs) of normally distributed variates. Normality of the groups were tested using the Shapiro-Wilk normality test with a type-I error threshold of α =0.05. When a group was found to be not normally distributed, outliers were removed and normality restored. Outliers correspond to values above or below 1.5 times the interquartile range (Tukey 1977).

The cut-off value estimated by Favre et al.'s (2008) method corresponds to the minimal joint prediction error risk to incorrectly attribute any individual value to one of the two groups, and thus offers the best compromise between the two prediction error risks. The farther the measured value is from the cut-off value separating the two groups, the lower is the error risk of assigning the corresponding specimen to the group located at the same side of the cut-off, leading to the computation of a prediction relative error risk - a quantity directly related to 'odds' as used in gambling. In horse racing the betting 'odds' expresses the amount of profit you will receive and the amount you have to bet to get it. For example, 1:5 (one fifth) or, similarly, the 'odds against' 5:1 (5 to 1), means you will get 5€ for every 1€ wagered. In the present case the prediction relative error risk can be expressed as odds written in the form of "r:s" (read: r sth, with r the betting and s the amount of profit) that corresponds to the probability of having a correct identification of p=s/(r+s). The relationship between odds and

probability can appear counterintuitive and complicated but only few values are important to remember. For instance, a betting odd of 1:100 (corresponding to the ratio of the probability that a prediction error is made to the probability that it is not made) will correspond to a probability $p = 100/101 \approx 99\%$ to correctly assign the specimen to the group located on the same side of the cut-off. Thus odds of 1:10 will correctly assign specimens to the group 90.9% of the time, odds of 1:5 to 83.3%, and odds of 1:2 to 66.7%.

Since a strong geographic variability exists in wild boar (e.g., Albarella et al., 2009; Rowley-Conwy et al 2012), analyses were carried out for the full dataset, as well as for all specimens from Europe (France, Switzerland, Germany and Poland) and for Eastern populations (Iran, Iran, Turkey, and Russia), separatly. All statistical analyses other than those computed with OpCut-Location v. 1.0 were performed using R v2.13.1 (R Development Core Team, 2011).

Identification of the specimens from the Durrington Walls

Specimens from the Durrington Walls were identified based on their molar lengths and widths according to the cut-off values and error risk curves established with the modern specimens. The evolution of the percentages of specimens correctly identified as domestic pigs was visualised according to the different threshold values outlined previously (1:100, 1:10, 1:5, 1:2 and cut-off values).

Results

Measurement values for both MTL-EMTL and MTW-EMTW are highly correlated (fig. 1), with coefficients of determination of 99.3% and 99.1%, respectively. For each least-square regression, the slope equals 1 (95% confidence intervals for the width: [0.972; 1.009], length: [0.975; 1.008]) and the intercept equals 0 (width: [-0.0176; 0.048], length: [-0.017; 0.114]), showing that the estimated measurements can be directly compared to the original ones without bias or loss of information. As a result, only EMTL and EMTW are used in the following analyses and designated subsequently as MTL and MTW for simplicity.

<Figure 1>

Statistics of the measurements used in the analyses are summarized in sup. table 1 (TSI-1) for the different wild boar populations, and in sup. table 2 (TSI-2) for the different domestic pig breeds. Summary of the statistics for wild boar and domestic pigs are reported in table 1.

Similar results obtained separately for European and Eastern populations are provided in sup. tables 3 and 4 (TSI-3/4).

<Table 1>

Not unexpectedly, for all measurements tested, domestic pigs appear to be significantly smaller than wild boars (table 1, fig. 2).

<Figure 2>

The cut-off values giving the optimal separation between the modern West Palearctic wild and domestic pigs were estimated for the four teeth and were as follows (table 2, fig. 3, 4):

- MTL: $2.39 \text{ cm for } M^2$, $3.69 \text{ cm for } M^3$, $2.26 \text{ cm for } M_2 \text{ and } 3.79 \text{ cm for } M_3$;
- MTW: 1.85 cm for M^2 , 2.13 cm for M^3 , 1.50 cm for M_2 and 1.75 cm for M_3 .

The corresponding relative error risk curves are represented in figures 3 and 4. The more the value is far from the cut-off value, the more confident is the identification. The risk of making a wrong identification does not decrease linearly and symmetrically on both sides of each cut-off value, and thresholds corresponding to the risks 1:100, 1:10, 1:5 and 1:2 are illustrated as dotted lines in figures 3 and 4. It is therefore possible to compare new values to the different thresholds, and to decide if a specimen can be identified or not according to levels of confidence. As expected, reliable identification can be obtained only for the extreme values (very small for domestic pigs, very large for wild boar). The higher the confidence, the fewer the number of definitively identified specimens – i.e. only 0.6% - 4.3% of the specimens identified with a risk of 1:100, 7.9% - 16.8% at 1:10, 17% - 28.8% at 1:5, and 50.9% - 62.6% at 1:2 (fig. 5), depending of the tooth and the measurement analysed.

For example, for the MTL of M₃, the cut-off value is 3.79 cm, meaning that specimens with an MTL lower than 3.79 cm are more likely to be domestic pigs. Conversely those showing an MTL higher than 3.79 cm more likely correspond to wild boar. Departing from this central cut-off value, a stricter threshold will increase the chance of providing the correct identification at the expense of the actual range of the variable that provides the actual identification. When the error risk limit is fixed to 1:10 (corresponding to a probability of correct assignment of ~90%), only the specimens with a MTL lower than 2.98 cm, or above 4.62 cm, could be identified as "domestic" or "wild" respectively. Between these two limit-

values is a zone of uncertainty where specimens cannot be safely identified as either wild or domestic under this 1:10 error risk constraint. Using a stricter threshold of 1:100 (~99%), only specimens below 2.57 cm and above 5.06 cm could be confidently assigned to their respective wild or domestic groups.

<Table 2>

<Figure 3>

<Figure 4>

<Figure 5>

Cut-off values (sup. tables 3 (TSI-3) and 4 (TSI-4)) and relative error risk curves (fig. sup. 1, 2, 3 and 4 (FSI-1/2/3/4)) obtained for Europe and the East (Near East and Russia) separately are slightly different, with cut-off and threshold values always smaller for Europe than for the Eastern populations (sup. tables 3 and 4 (TSI-3/4)).

Measurements from the UK Neolithic pig standard of Durrington Walls (Albarella & Payne 2005) were identified by our method as domestic in 92% to 100% of the cases, depending on the measurement used and the tooth analysed (fig. 6). Only the MTL of the M₂ provided a lower success rate, with only 79.5% of the specimens identified as domestic. All the remaining specimens were identified as wild boar with a high error risk (17 measurements between the cut-off value and the 1:2 threshold, and one between 1:2 and 1:5). Pooling all the analyses together, 94% of the teeth were identified as domestic pigs. In this analysis, again, there are fewer specimens identified with low error risk than specimens identified with measurements close to the cut-off values (fig. 6).

<Figure 6>

Discussion

Differentiating wild from domestic forms of mammals and birds has been a major focus of zooarchaeological research for decades (e.g. Vigne et al. 2005) – particularly those associated with the transition from hunting to herding. Our understanding of this major human biocultural transition relies upon our ability to explore the domestication process itself in more detail and to develop more robust tools with which to achieve that. Pigs have received particular attention in this respect over the last years (e.g., Albarella et al. 2005, 2006,

Rowley-Conwy et al. 2012 for the most recent). Wild and domestic forms have been traditionally separated using a measure of size, especially on the third lower molar length and width (Payne and Bull 1988, Albarella et al. 2006, Rowley-Conwy et al. 2012). The present study provides a new more extensive baseline dataset of modern comparative dental size measurements, and a more robust and rigorous statistical tool for use in identifying the domestic or wild status of *Sus scrofa* remains from archaeological sites.

For the maximum tooth length measurements, the cut-off values established are 2.39 cm for M², 3.69 cm for M³, 2.26 cm for M₂, 3.79 cm for M₃, whereas for maximum tooth width they are respectively 1.85 cm, 2.13 cm, 1.50 cm, and 1.75 cm respectively. Specimens with values below these cut-offs more likely correspond to domestic pigs, and above to wild boars. The cut-off values for maximum tooth length and width were measured and associated with curves and threshold values estimated for relative error risks of 1:2, 1:5, 1:10 and 1:100, corresponding to prediction of correct probabilities of ~67%, ~83%, ~91% and 99%, respectively. Using the threshold of 1:10, around 10% of the analyzed specimens could be correctly identified with a probability of ~91%. A threshold of 1:100 will raise the confidence of identification significantly; however, a high proportion of the specimens (>95%) will remain unidentified. A threshold of 1:2 will allow ~55% of the specimens to be identified, but with a probability of correct identification of only 66%. Whilst each individual researcher must decide on the level of acceptable error, obviously based upon the specific archaeological questions under scrutiny, this approach at least provides some basic quantitative data informing how identifications have been made. However, what remains clear from the above results is that linear cheek tooth dimensions offer extremely low power in discriminating between wild and domestic Sus scrofa specimens.

The method of establishing cut-off values for identifying the wild or domestic status of archaeological pig remains goes back to the roots of the discipline of archaeozoology. Rütimeyer (1860) first published ranges of measurements for prehistoric wild boar (*Sus scrofa ferus*), with M₃ lengths ranging from 4.0 to 5.3 cm, and for a smaller 'domestic' group (called *Sus scrofa palustris*) ranging between 3.3 and 3.9 cm. This latest group would be considered today as domestic. In a more recent study, Mayer et al. (1998) identified threshold values for maximum tooth length and width for hybrids and feral pigs used as surrogates for domestic pigs, and minimum values for wild boar. All our width threshold values published here fall within their intervals, whereas only our M₃ length threshold is included in their interval, with

all other length thresholds we obtained being slightly larger than the ones of Mayer et al. (1998). These incongruences may be due to the measurement techniques used by Mayer et al. (measurements were taken using dial callipers), the geographic origin of the samples, or (more likely) by the fact that they used hybrids and feral pigs instead of true domestic pigs. Comparing the full range of variability within each group is perhaps not as relevant for local studies as it might be for broader temporal and geographic syntheses. Indeed, recent and extant wild boar shows a large variability of size across its full Old world range (e.g. Albarella et al. 2009), as do the different modern domestic breeds (e.g. Schaaf, 1953). Cut-off and threshold values obtained for European (France, Switzerland, Germany, Poland) and Eastern (Iran, Iraq, Turkey and Russia) wild boar are only slightly different, with Eastern populations always presenting larger values than those from Europe. All the measurements provided in supplementary data can be used to perform similar computations based on even more restricted geographic subsamples, or using only subsets of domestic pigs.

The vast majority (94%) of the specimens from the Late Neolithic site of Durrington Walls (Albarella & Payne 2005) were identified as domestic pigs based on our cut-of values. According to Bull and Payne (1982), data based on highly improved modern pig breeds should not be used to interpret archaeological data due to their reduced relevance to wild boar or ancient domestic breeds. Because the cut-off values and error risk curves provided in this study have been established using recent specimens, further comparisons with ancient specimens of known status are required before generalising to the zooarchaeological record. Modern 'wild' and 'domestic' pigs are the two extremes of a domestication continuum. Archaeological records evidenced that pigs have gradually and slowly changed during the domestication process (Ervynck et al. 2001), potentially resulting in changes of the cut-off values and error risk curves through time.

Recent genetic and morphometric evidence for the introduction of domestic pigs of Near eastern/Anatolian origin to Europe during the Neolithic, followed by the subsequent rapid incorporation of European wild boar lineages into domestic swineherds (Larson et al 2007; Otonni et al 2013) must mean that both large, small and intermediate-sized domestic pigs should be expected on Neolithic archaeological sites across Europe. As a note of caution, a recent study, involving a combined aDNA and Geometric morphometric approach, has indeed revealed the presence of large but clearly domestic pigs at early Linearbandkeramik and late

Ertebølle sites in northern Germany (Krause-Kyora et al. in press), contradicting the traditionally accepted view that domestic pigs are small, and wild boar are large.

According to the present study and the results obtained by Evin et al. (2013), size is a less than ideal criteria to identify modern wild and domestic Sus specimens, since the majority of specimens fall close to the cut-off values and can only be identified with a high error risk. Nevertheless, size is often one of the only variables currently available in the published literature. By its capacity to include finer differences, shape (NOT size) remains the most powerful descriptor to identify wild boar and domestic pigs based on their dentition, and should be the first choice over the traditional biometrical techniques where discrimination is the principal research question (Evin et al. 2013). It is often quicker to measure linear distances than to acquire geometric data that require preliminary handling and treatments before the computation of statistical analyses. The effort and time required by both techniques has therefore to be considered in terms of the balance between questions asked and level of information required. It is clear that in terms of wider comparative zooarchaeological information, linear measurement datasets currently vastly outnumber those using Geometric Morphometric approaches. As a result, it is important to utilise these exiting datasets in new more systematic and quantitative ways, whilst at the same time being aware of and highlighting their limitations.

Conclusion

This study provides a new biometric framework for distinguishing modern West Palearctic wild and domestic pigs that can be applied to existing biometrical datasets that incorporate linear measurements of molar teeth. The statistical tool presented in this study provides cut-off values paired with error risks and therefore offers more objective criteria for identifying wild and domestic pigs using simple measurements of maximum tooth length and width. This framework provides a much more extensive sampling of both modern wild boar populations and domestic breeds than has so far been available. The quantification of the error risk related to identification will allow researchers to revisit past and present biometrical datasets in order to more systematically assess the likley presence and proportions of wild and domestic *Sus scrofa* represented.

Domestication is a continuous and ongoing process, therefore providing cut-off values can be seen as somewhat artificial. However, the continuity of the process is reflected by the

continuity of size. Accordingly, there is a risk that a relatively large number of specimens will have measurements close to the cut-off values and therefore should be kept unidentified. We recommend the use of the cut-off and threshold values presented in this study only when wider comparative analyses is required with other biomerical datasets and where more powerfull analyses of shape are not available.

References

- Albarella, U., Payne, S., 2005. Neolithic pigs from Durrington Walls, Wiltshire, England: a biometrical database. J. Archaeol. Sci. 32, 589-599.
- 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. and Smith, B.D. (Eds), Documenting Domestication: New Genetic and Archaeological Paradigms. University of California Press, Berkeley, pp. 209e227.
- 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. Environmental Archaeology, 14, 103–136.
- Aulagnier, S., Haffner, P., Mitchell-Jones, T., Moutou, F., Zima, J. 2008. Guide des mammifères d'Europe, d'Afrique du Nord et du Moyen-Orient. Paris: Delachaux et Niestlé SA.
- Bull, G., Payne, S., 1982. Tooth eruption and epiphysial 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 109, pp. 55-71.
- Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A., 1983. Graphical Methods for Data Analysis. Wadsworth & Brooks/Cole.
- 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., Hulme-Beaman, a., Yuan, J., Dobney, K., 2011. Early Neolithic pig domestication at Jiahu, Henan Province, China: clues from molar shape analyses using geometric morphometric approaches. J. Archaeol. Sci. 38, 11-22.
- Darwin, C. 1868. The variation of animals and plants under domestication. John Murray, London.
- Evin, A., Cucchi, T., Cardini, A., Strand Vidarsdottir, U., Larson, G., Dobney, K., 2013. The long and winding road: identifying pig domestication through molar size and shape. J. Archaeol. Sci. 40, 735–743.

- 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, 47e73.
- Favre, E., Escarguel, G., Suc, J., Vidal, G., & Thévenod, L., 2008. A contribution to deciphering the meaning of AP / NAP with respect to vegetation cover. Rev. Palaeobot. Palyno., 148, 13–35.
- 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.
- Krause-Kyora, B., Makarewicz, C., Evin, A., Girdland-Flink, L., Dobney, K., Larson, G., Hartz, S., Schreiber, S., von Carnap-Bornheim, C., von Wurmb-Schwark, N., Nebel, A. in press. Use of domesticated pigs by Mesolithic hunter-gatherers in northwest Europe. Nat. Commun.
- Larson, G., Dobney, Keith, 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, 1618-21.
- Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J-D., Edwards, C. J., Schlumbaum, A., Dinu, A., Balasescu, A., Dolman, G., Tagliacozzo, A., Manaseryan N., Miracle, P., van Wijngaarden-Bakker, L., Masseti, M., Bradley, D. G., Cooper, A. and 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. Proc. Natl. Acad. Sci. U.S.A. 104, 15276–15281.
- Larson, G., Cucchi, T., Fujita, Masakatsu, 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., Bergh, G. Van den, West, E., Wickler, S., Yuan, Jing, 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, 4834-9.
- Larson, G., & Burger, J. 2013. A population genetics view of animal domestication. Trends Genet. 29, 197–205
- Ludwig, A., Pruvost, M., Reissmann, M., Benecke, N., Brockmann, G. A., Castaños, P., Cieslak, M., Lippold S., Llorente L., Malaspinas A.-S., Slatkin M., Hofreiter M. 2009. Coat Color Variation at the Beginning of Horse Domestication. Science 324, 485.
- 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.
- 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.
- Ottoni, C., Girdland Flink, L., Evin, A., Geörg, C., De Cupere, B., Van Neer, W., Bartosiewicz, L., et al. 2013. Pig Domestication and Human-Mediated Dispersal in Western Eurasia Revealed through Ancient DNA and Geometric Morphometrics. Mol. Biol. Evol. 30, 824-832.
- Payne, S., Bull, G., 1988. Components of variation in measurements of pig bones and teeth, and the use of measurments to distinguish wild from domestc pig remains. Archaeozoologia II, 27-66.

- R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rohlf, F.J., 2010. tpsDig. Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY.
- 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.
- Rütimeyer, L., 1862. Die Fauna der Pfahlbauten der Schweiz. Bürkli, Basel.
- Schaaf A. 1953. Beziehungen zwischen Körper-, Skelett- und Schädelmaßen und dem Ansatz von Fleisch und Fett beim Schwein Wissenschaftliche Abhandlungen, Band II. Deutsche Akademie der Landwirtschaft zu Berlin, Leipzig.
- Tukey, J. W., 1977. Exploratory Data Analysis, Addison-Wesley.
- Trut, L.N., 1999. Early canid domestication: the farm-fox experiment. Am. Sci. 87, 160–169.
- von den Driesch, A., 1976. A Guide to the Measurement of Animal Bones from Archaeological Sites. Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, Mass.
- Vigne, J.-D., Peters, J., Helmer, D., 2005. The First Steps of Animal Domestication: New Archaeozoological Techniques. Oxbow Books, Oxford.

 Figure 1: Linear relation and correlation between (a) Maximum Tooth Length (MTL) and Estimated Maximum Tooth Length (EMTL) and (b) Maximum Tooth Width (MTW) and Estimated Maximum Tooth Width (EMTW) measured on lower M3. Results of the regression tests are provided as adjusted \mathbb{R}^2 and associated p-value.

Figure 2: Boxplot visualization of Maximum Tooth Length (MTL) and Width (MTW) variability between domestic pigs (DP) and wild boars (WB). Length and width are expressed in cm.

Figure 3: Relative risk curves for erroneous "wild" versus "domestic" prediction based on Maximum Tooth Length (MTL). The curve to the right of the cut-off value corresponds to the relative risk of falsely assigning the specimen to a wild boar, while the curve to the left corresponds to the relative risk of falsely assigning the specimen to a domestic pig. Critical values for which the prediction relative error risk is 1:100 (~99%), 1:10 (~91%), 1:5 (~83%) and 1:2 (~67%) are shown at the intersection between the corresponding relative risk value and risk curve. At the top are the numbers of specimens of wild (nWB) or domestic (nDP) pigs corresponding to the range delimited by the critical values.

Figure 4: Relative risk curves for erroneous "wild" versus "domestic" prediction based on Maximum Tooth Width (MTW). See Fig. 3 for explanations.

Figure 5: Percentage of specimens identified for each relative error risk threshold. From left to right: the first four values correspond to MTL (upper M2 and M3, lower M2, and M3) and the four last to MTW (same order).

Figure 6: Percentage of Durrington Walls specimens identified as domestic pigs for each relative error risk threshold.

Table 1: Sample size, mean, standard deviation (sd), minimal (min) and maximal (max) values for wild boars and domestic pigs for upper and lower second (M2) and third (M3) molars for maximum tooth length (MTL) and width (MTW), as well as results (W and *p*-value) of Wilcoxon's tests for differences between the two groups. * shows where outliers have been removed.

Table 2: Cut-off values at some critical threshold values for relative error risks of erroneous identification based on maximum tooth length (MTL) and width (MTW). The four odds retained here, 1:100, 1:10, 1:5 and 1:2, correspond to probabilities of wrong assignment of ~99%, ~91%, ~83% and ~67%, respectively.

Supplementary information:

Table TSI-1: Sample size, mean, standard deviation (sd), median (Med.), minimal (min) and maximal (max) measurement values for wild boar divided into country of origin for upper and lower second (M2) and third (M3) molars for maximum tooth length (MTL) and width (MTW).

Table TSI-2: Sample size, mean, standard deviation (sd), median (Med.), minimal (min) and maximal (max) measurement values for domestic pigs divided into breeds for upper and lower second (M2) and third (M3) molars for maximum tooth length (MTL) and width (MTW).

Table TSI-3: Sample size, mean, standard deviation (sd), median (Med.), minimal (min) and maximal (max) measurement values for European (France, Switzerland, Germany, Poland) and Eastern (Iran, Iraq, Turkey, Russia) wild boar populations for upper and lower second (M2) and third (M3) molars for maximum tooth length (MTL) and width (MTW).

Table TSI-4: Cut-off values and critical threshold values for relative error risks of erroneous identification of domestic pigs (DP) and European wild boar (WB) (A); or Eastern Wild boar (B) based on maximum tooth length (MTL) and width (MTW), at the odds 1:100, 1:10, 1:5 and 1:2, corresponding to probabilities of wrong assignment of ~99%, ~91%, ~83% and ~67%, respectively. Corresponding error curves are in sup. Figures FSI-1 to FSI-4. The "European" group contains specimens from France, Switzerland, Germany and Poland; the "Eastern" group contains specimens from Iran, Iraq, Turkey and Russia.

Figure FSI-1: Relative risk curves for erroneous "European wild" versus "domestic" identification based on Maximum Tooth Length (MTL). The curve to the right of the cut-off value corresponds to the relative risk of falsely assigning the specimen to a wild boar, while the curve to the left corresponds to the relative risk of falsely assigning the specimen to a

domestic pig. The "European wild" group contains specimens from France, Switzerland, Germany, and Poland.

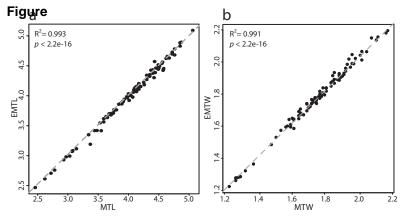
Figure FSI-2: Relative risk curves for erroneous "European wild" versus "domestic" identification based on Maximum Tooth Width (MTW). The curve to the right of the cut-off value corresponds to the relative risk of falsely assigning the specimen to a wild boar, while the curve to the left corresponds to the relative risk of falsely assigning the specimen to a domestic pig. The "European wild" group contains specimens from France, Switzerland, Germany, and Poland.

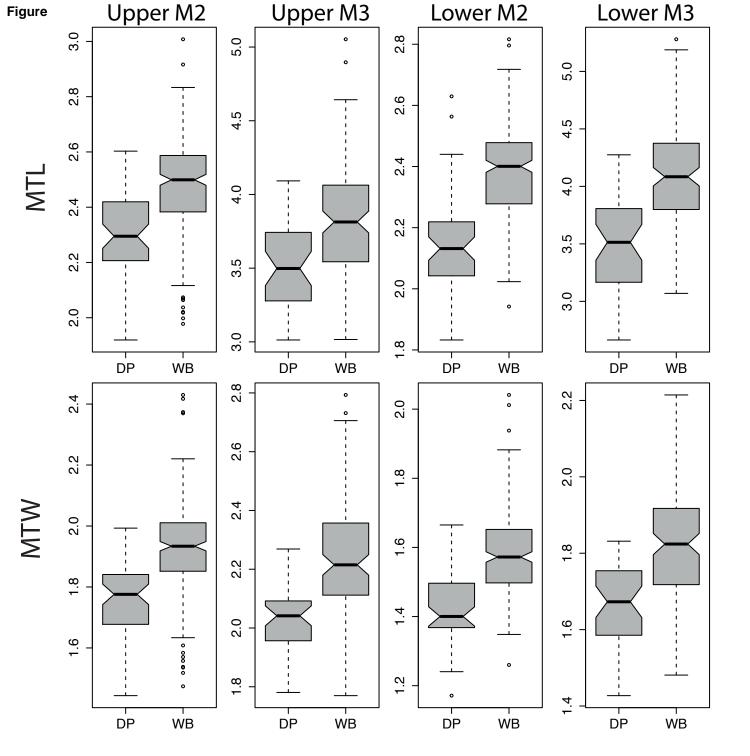
Figure FSI-3: Relative risk curves for erroneous "Eastern wild" versus "domestic" identification based on Maximum Tooth Length (MTL). The curve to the right of the cut-off value corresponds to the relative risk of falsely assigning the specimen to a wild boar, while the curve to the left corresponds to the relative risk of falsely assigning the specimen as domestic pig. The "Eastern wild" group contains specimens from Iran, Iraq, Turkey, and Russia.

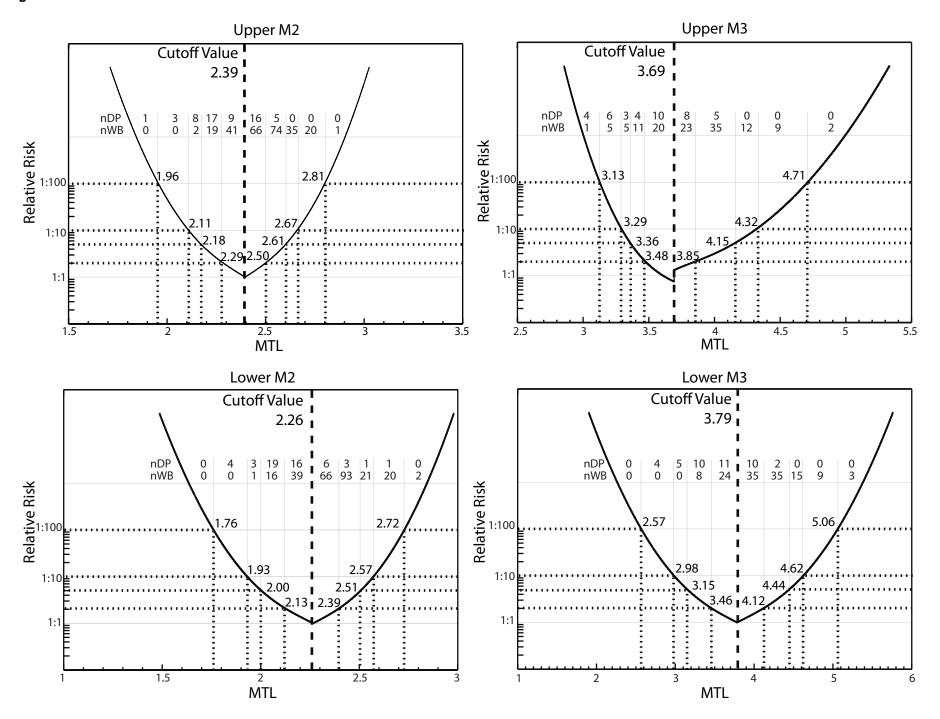
Figure FSI-4: Relative risk curves for erroneous "Eastern wild" versus "domestic" identification based on Maximum Tooth Width (MTW). The curve to the right of the cut-off value corresponds to the relative risk of falsely assigning the specimen to a wild boar, while the curve to the left corresponds to the relative risk of falsely assigning the specimen to a domestic pig. The "Eastern wild" group contains specimens from Iran, Iraq, Turkey, and Russia.

		Wild Boar				Domestic Pig				Differences			
		N	mean	sd	min	max	N	mean	sd	min	max	W	р
MTL	Upper M2	258*	2.490*	0.137*	2.116*	2.833*	59	2.301	0.148	1.92	2.603	12611	8.30E-13
	Upper M3	123	3.84	0.373	3.016	5.053	40	3.51	0.164	3.013	4.092	3717	1.26E-06
	Lower M2	258	2.379	0.15	1.942	2.816	53	2.142	0.164	1.833	2.63	11741	< 2.2e-16
	Lower M3	129	4.085	0.419	3.069	5.28	42	3.496	0.399	2.664	4.275	4568	2.57E-11
MTW	Upper M2	257*	1.931*	0.112*	1.608*	2.220*	59	1.762	0.119	1.444	1.993	13122	2.13E-15
	Upper M3	123	2.23	0.177	1.77	2.793	40	2.026	0.1	1.781	2.269	4190	2.53E-11
	Lower M2	254*	1.575*	0.107*	1.348*	1.882*	53	1.426	0.104	1.171	1.665	11475	7.39E-15
	Lower M3	129	1.823	0.146	1.481	2.214	42	1.661	0.108	1.427	1.831	4389	1.67E-09

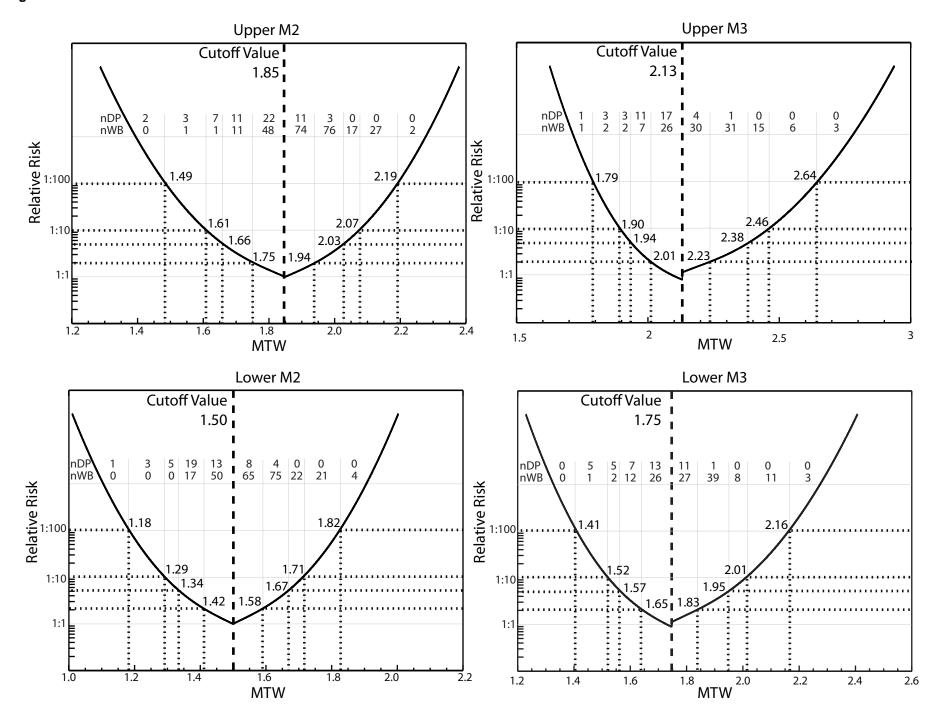
		Cut-off		Risk to p	redict DP		Risk to predict WB				
			1:100	1:10	1:5	1:2	1:2	1:5	1:10	1:100	
MTL	Upper M2	2.39	1.96	2.11	2.18	2.29	2.50	2.61	2.67	2.81	
	Upper M3	3.69	3.13	3.29	3.36	3.48	3.85	4.15	4.32	4.71	
	Lower M2	2.26	1.76	1.93	2.00	2.13	2.39	2.51	2.57	2.73	
	Lower M3	3.79	2.57	2.98	3.16	3.46	4.12	4.44	4.62	5.06	
MTW	Upper M2	1.85	1.49	1.61	1.66	1.75	1.94	2.03	2.07	2.19	
	Upper M3	2.13	1.79	1.90	1.94	2.01	2.23	2.38	2.46	2.64	
	Lower M2	1.50	1.18	1.29	1.34	1.42	1.58	1.67	1.71	1.82	
	Lower M3	1.75	1.41	1.52	1.57	1.65	1.83	1.95	2.01	2.16	

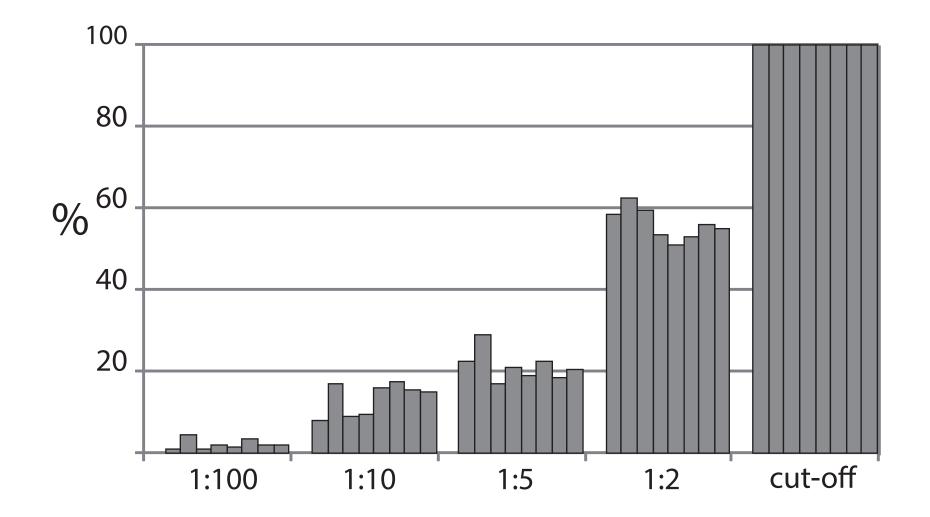


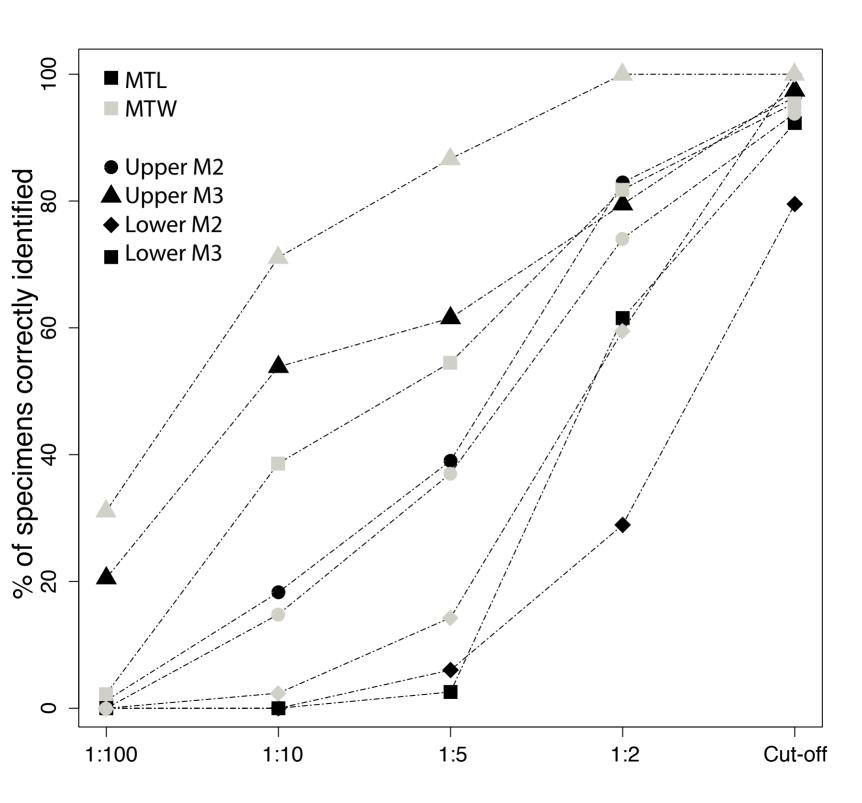




Figure







Supplementary Material
Click here to download Supplementary Material: fig-SI1.pdf

Supplementary Material
Click here to download Supplementary Material: fig-SI2.pdf

Supplementary Material
Click here to download Supplementary Material: fig-Sl3.pdf

Supplementary Material
Click here to download Supplementary Material: fig-SI4.pdf

Supplementary Material
Click here to download Supplementary Material: Sup-tables.xlsx