A landmark-based approach for assessing the reliability of mandibular tooth crowding as a marker of dog domestication

Carly Ameen [a, *, 1], Ardern Hulme-Beaman [a, b, *, 1], Allowen Evin [c, a], Mietje Germonpré [d], Kate Britton [e, f], Thomas Cucchi [g, e], Greger Larson [b], Keith Dobney [a, e, h]

[a] Department of Archaeology, Classics and Egyptology, University of Liverpool, 12-14 Abercromby Square, Liverpool, L69 7WZ, UK
[b] Palaeogenomics & Bio-Archaeology Research Network, Dyson Perrins Building, South Parks Road, Oxford, OX1 3QY, UK
[c] Institut des Sciences de l'Evolution, Université de Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095, Montpellier Cedex 05, France
[d] Operational Direction "Earth and History of Life", Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000, Brussel, Belgium
[e] Department of Archaeology, University of Aberdeen, St. Mary's Building, Elphinstone Road, Aberdeen, UK
[f] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103, Leipzig, Germany
[g] CNRS-Muséum National d'Histoire Naturelle, UMR 7209, Archéozoologie, Archéobotanique Sociétés, Pratiques et Environnement, 55 rue Buffon, 75005 Paris, France
[h] Department of Archaeology, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada

ARTICLE INFO

ABSTRACT

Tooth crowding is one of several criteria used to infer the process of domestication in the zooarchaeological record. It has been primarily used to support claims of early animal domestication, perhaps most contentiously in claims for the existence of so-called "proto-domestic" dogs as early as the Middle-Upper Palaeolithic. Tooth crowding studies vary in their methodological approaches, and interpretation of the resulting data is constrained by the limited geographic and temporal scope of reference specimens used to construct an appropriate comparative framework. To address these key problems, we present a standardised landmark-based protocol for the measurement and quantification of mandibular tooth crowding that can be systematically applied in the context of dog domestication research. We then test the assumption that tooth crowding is less frequent in ancient and modern wild wolf populations by examining 750 modern dogs and 205 modern wolves from across the modern geographic range of Canis lupus as well as 66 Late Pleistocene wolves from Alaska.

Our results demonstrate that landmark-based metrics provide a reliable approach for recording and analysing tooth crowding. Although it is likely that the relatively low frequency of tooth crowding found in our modern dog dataset (~6%) in part reflects the 'modern' morphology of domestic breeds, the higher frequency of crowding in both modern (~18%) and ancient (~36%) wolves strongly suggests that current assumptions linking tooth crowding with the process of early domestication (at least in dogs) should be critically re-evaluated, and that further investigations into the drivers behind these developmental patterns should be pursued.

1. Introduction

A suite of phenotypic changes associated with the domestication process in mammals has been observed (and studied) in virtually all animal domesticates (e.g. Darwin, 1868; Clutton-Brock, 1999; Zeder, 2012). Traditionally, morphological changes to the skull (e.g. snout shortening, cranial flexion, and tooth size reduction; see Wayne, 1986; Clutton-Brock, 1999; Morey, 1992; Drake, 2011; Zeder, 2012), as well as size reduction of elements of the appendicular skeleton, have been the principal signature with which to track domestication in the zooarchaeological record. Another regularly accepted criterion is the presence of tooth crowding, where tooth orientation and alignment is described as touching, overlapping and/or rotated. Although there is no universally accepted definition for what constitutes a crowded toothrow, it is traditionally considered an important

characteristic of initial domestication in dogs, (e.g. Lawrence, 1967; Clutton-Brock, 1963, 1999; Benecke, 1987; Morey, 1992; Moray, 1994; Germonpré et al., 2012, 2015a,b), and other domesticates (e.g. pigs; Krause-Kyora et al., 2013) since it is often linked with snout shortening found commonly in many domestic species. The prevalence of tooth crowding has been used as a specific criterion to identify dogs in the archaeological record (e.g. Degerbøl, 1961; Benecke, 1987; Dimitrijević and Vuković, 2012), and to support claims for the existence of Middle-Upper Palaeolithic dogs (Germonpré et al., 2012, 2015a, b; Ovodov et al., 2011).

The identification of so-called 'Palaeolithic dogs' is controversial, and researchers have suggested that additional evidence is required to support such claims (Crockford and Kuzman, 2012; Germonpré et al., 2013; Morey, 2014; Boudadi-Maligne and Escarguel, 2014; Germonpré et al., 2015a). Studies have also acknowledged that tooth crowding alone is insufficient to justify claims for the presence of early dogs (e.g. Davis and Valla, 1978; Sablin and Khlopachev, 2002; Ovodov et al., 2011). This is especially pertinent since tooth crowding has also been observed in both wild and captive wolf populations (e.g. Degerbøl, 1961; Lawrence, 1967; van Wijngaarden-Bakker, 1974; Dimitrijević and Vuković, 2012), and the natural variation of Late Pleistocene wolves remains poorly documented and not well understood (Larson et al., 2012; Crockford and Kuzman, 2012; Perri, 2016). Thus, despite its ubiquity as a proxy for domestic status, the link between tooth crowding and early domestication processes remains tenuous.

1.1. Previous methods and analyses

Previous methods for studying tooth crowding vary. In some studies, the occurrence of crowding is recorded through a visual assessment of overlapping teeth in the upper and/or lower jaw (Sablin and Khlopachev, 2002; Germonpré et al., 2012, 2015a,b; Napierala and Uerpmann, 2012). Others have developed simple biometric protocols for recording crowding using a series of ratios from length measurements of the molars and premolars (e.g. Lawrence, 1967; Davis and Valla, 1978; Benecke, 1987; Musil, 2000; Lapham, 2010; Dimitrijević and Vuković, 2012). Degerbøl (1961: 39) applied a more systematic method for measuring crowding of the maxillary toothrow as a ratio of the cumulative length of the three anterior premolars (P1, P2, P3) and the length between the canine (C1) and the carnassial (P4). Clutton-Brock (1963) later expanded this analysis to the mandible, using a ratio of the sum of the lengths of all teeth between P2 and the M3, against the total length of the toothrow measured from P2 to M3.

Degerbøl (1961) and Clutton-Brock (1963) protocols were adopted by others to assess the domestic status of canid remains from the Neolithic sites of Newgrange (Ireland: van Wijngaarden-Bakker, 1974:342), and Staines and Road Farm (England: Clark, 1996). In the case of Staines and Road Farm, measurements were limited to lower premolars only, and a new formula was defined for the maxilla ("Lengths P1 + P2 + P3) X 100/length anterior edge P1 to anterior edge P2" — Clark 1996: 214), which differs significantly from others in terms of defining the toothrow. However, since the reported indices are largely in line with previously reported ranges (Clark, 1996: 214, table 2), it can be assumed that this definition is incorrect, and that in fact the maxillary toothrow was measured to the anterior edge of the P4 (not the P2), following Degerbøl (1961), and Clutton-Brock (1963), although only van Wijngaarden-Bakker (1974) is referenced.

These studies have produced tooth-crowding indices of continuous data that are then directly compared with other datasets, where indices of tooth crowding in wild and domestic specimens have previously been calculated (e.g. Clark, 1996; van Wijngaarden-Bakker, 1974; Walker and Frison, 1982; Ovodov et al., 2011). For instance, Ovodov et al. (2011) measured the tooth-crowding index of a 33,000-year-old canid from Razboinichya Cave (Russia) and compared it to crowding indices reported from Clark (1996) Neolithic dog samples, as well as Benecke (1994) Předmostí (Czech Republic) canid mandibles, but not with other contemporaneous specimens. The authors do, however, urge caution when using only tooth crowding as evidence for the possible presence of early domesticate dogs (Ovodov et al., 2011).

The reference datasets used as a baseline for tooth crowding in wild wolf populations are often limited both in number of specimens and geographic coverage (e.g. only modern European wolves are used in Clark, 1996). Inconsistency of the methods and metrics, as well as the development of study specific measurements (i.e. Dimitrijević and Vuković, 2012), also hinders comparisons with previously measured archaeological and wild canid specimens. Statistical analyses are additionally lacking in previous studies of tooth crowding. Although Benecke (1987) and Dimitrijević and Vuković (2012) use discriminate analysis on mandibular and maxilla measurements to study Upper Palaeolithic canids from Northern Europe and Mesolithic/Early Neolithic dogs from the Danube Gorge respectively, neither attempted to discriminate levels of tooth crowding between wild and domestic animals, even though both studies continue to note the importance of tooth crowding as an indicator of domestication (Benecke, 1987:33; Dimitrijević and Vuković, 2012).

As a result, there remains both a methodological and contextual disconnect between the recording and the interpretation of tooth crowding data. The lack of a systematic recording protocol and associated statistical methods for quantifying tooth crowding among dogs and wolves means the overriding hypothesis that crowding is a product of domestication has yet to be empirically tested on appropriate modern wild and domestic comparative material.

To address this, we refined and adapted the methods of Degerbøl (1961) and Clutton-Brock (1963) for recording mandibular tooth crowding using easily applicable landmark-based approaches. We then applied these protocols to a large sample of modern/recent domestic dog and wolf mandibles, along with a sample of Pleistocene wolves. We first tested which (if any) group showed the highest proportion of specimens with tooth crowding, as well as which group contained specimens with the highest overall crowding value. We then tested whether instances of tooth crowding differed between wolves and dogs, and whether it was possible from these data to distinguish wolves and dogs based on these measures of tooth crowding.

2. Materials

A total of 1021 specimens were analysed, including 750 modern domestic dogs, 205 modern Grey wolves (Canis lupus), and 66 Pleistocene wolves. The modern dog sample derived from the collection at the Natural History Museum of Bern (Switzerland) and consists of pure bred individuals collected or donated to the museum since the early 20th century. The late Pleistocene wolves are all from Alaska, USA (housed in the American Museum of Natural History), 14 of which have been directly dated to between 45,500 ± 2700 uncal BP to 15,268 ± 169 uncal BP (Leonard et al., 2007 & Supplementary Information; also SI 2, SI Fig. 2 this paper). The modern wolf specimens cover the full extent of Canis lupus' natural range in both North America and Eurasia (Fig. 1). These modern wolves came from collections held at the Smithsonian Institute (Washington DC), Lisbon Natural History Museum,

Institute of Zoology Chinese Academy of Sciences (Beijing), Harvard Museum of Comparative Zoology (Cambridge, MA), Naturalis Museum (Leiden) and the British Natural History Museum (London).

Modern dog breeds were first analysed as a complete group, then two subgroups consisting of so-called wolf-like breeds ($n = 188$), identified as such by their retention or exhibition of more wolf-like characteristics (Table 1; see also Discussion 5.4), and Pariah dogs ($n = 65$), consisting of village or feral dogs with good provenance were analysed separately. All specimens were anatomically adult exhibiting fully erupted dentitions. Sex was recorded for 636 dogs (335 males, 301 females) and 108 modern wolves (56 males, 52 females), and it was possible to assess if sexual dimorphism influenced the pattern of tooth crowding in both groups using these data.

Coordinates of fourteen landmarks were recorded on 2D digital photographs of the occlusal surface of the hemi-mandible with all premolars and molars visible. Photographs were taken using a Nikon D5100 DSLR camera fitted with a Nikkor 60 mm AF-S micro lens. The placement of the hemi-mandible was controlled by adjustment of the relative position of the first molar. Landmarks were digitized using TPSDig2 v2.17 (Rohlf, 2013), and were placed at the anterior and posterior extremes of the longest axis of each mandibular premolar and molar (or alveoli if the tooth was absent) (Fig. 2). These positions define the greatest length of each tooth, regardless of position or rotation along the mesial/distal axis (SI Fig. 1).

3. Methods

3.1. The measurements

3.1.1. Establishing toothrow length

We defined two methods for quantifying overall toothrow length (Fig. 2a and b). The first is the maximum toothrow length (MTL), defined as the linear distance between landmark 1 and landmark 14 (Fig. 2a). The second is the cumulative toothrow length (CTL), and measures the cumulative distance from the posterior landmark of the third molar (landmark 1) through the anterior landmark of all teeth in sequence, ending with the anterior landmark of the first premolar (landmark 14) (Fig. 2b).

Table 1
List of breeds included in the wolf-like group used in this study, with sample size and source of reference for classification as wolf-like.

Wolf-like Breeds	Sample size	Source
Akita Inu	8 (3 female/3 male)	Parker et al., 2004
Alaskan Malamute	4 (1 female/3 male)	Parker et al., 2004
Chow Chow	15 (9 female/5 male)	Parker et al., 2004
Eskimo Dog	9 (0 female/4 male)	Leonard et al., 2002
German Shepherd	63 (20 female/42 male)	Germonpré et al., 2009
Great Dane	29 (18 female/9 male)	Germonpré et al., 2009
Greenland Dog	10 (4 female/6 male)	Leonard et al., 2002
Irish Wolfhound	21 (12 female/7 male)	Germonpré et al., 2009
Mastiff	2 (0 female/2 male)	Germonpré et al., 2009
Samoyed	1 (0 female/1 male)	Parker et al., 2004
Shar Pei	2 (2 female/0 male)	Parker et al., 2004
Siberian Husky	19 (13 female/6 male)	Parker et al., 2004
Tibetan Mastiff	5 (2 female/2 male)	Germonpré et al., 2009
Total	**188**	

3.1.2. Protocols for measuring crowding

To quantify tooth crowding in the mandible, we developed the following two formulas:

The Spacing Ratio (SR) calculates the difference between the mandibular toothrow length (TRL) and the cumulative length of all mandibular teeth. It is a ratio between the sum of the length of all mandibular teeth (CMM) (Fig. 2c) and TRL, which can be measured by either the cumulative toothrow length (CTL, Fig. 2b) or the maximum toothrow length (MTL, Fig. 2a)(Equation (1)).

$$SR = \frac{CMM}{TRL} \tag{1}$$

The Rotation Ratio (RR) calculates a ratio whose numerator is the sum of the difference between CMM and TRL, plus the difference between the cumulative distance between all 14 landmarks consecutively (ZZ metric) (Fig. 2d) and TRL, and whose denominator is TRL (Equation (2)). This metric incorporates alignment and rotational changes to the mandibular teeth as a result of crowding.
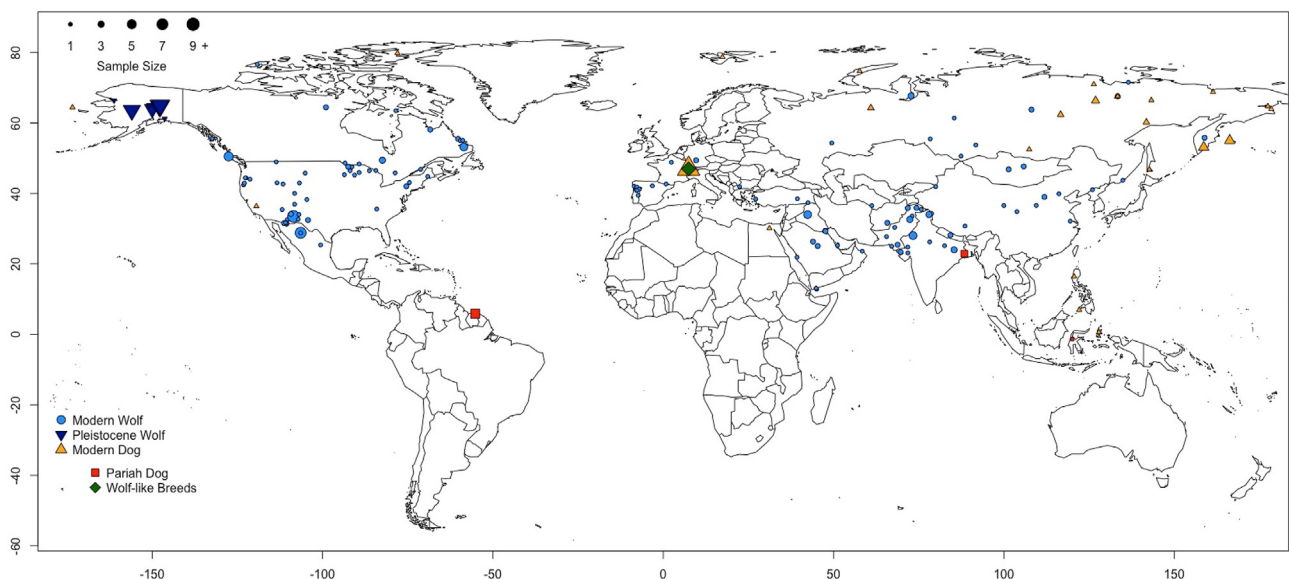
$$RR = \frac{(CMM - TRL) + (ZZ - TRL)}{TRL} \tag{2}$$

Fig. 1. Map showing sample size and collection location of specimens used in this study.
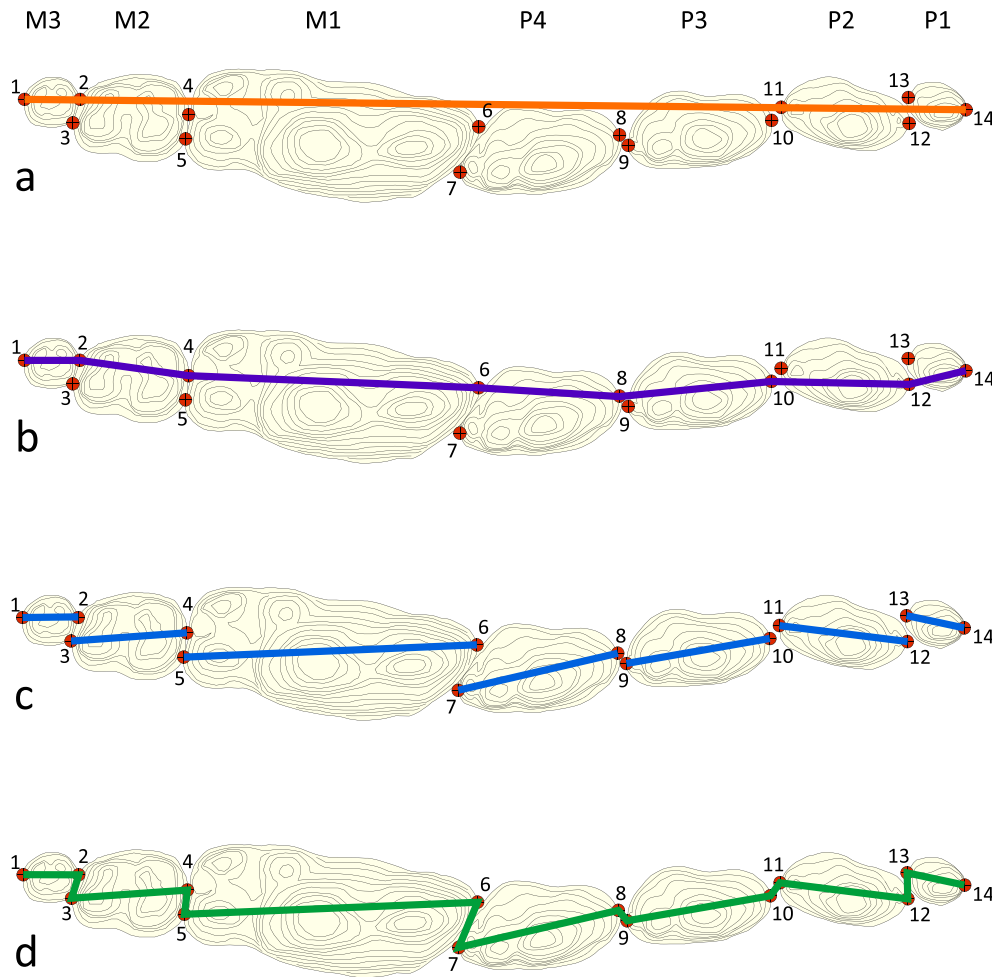
Fig. 2. Diagrams of the mandibular toothrow with the location of the 14 landmarks used in the study and visualisation of (a) the traditional, maximum toothrow length (MTL), (b) the cumulative toothrow length (CTL), (c) the cumulative length (CMM) metric, and (d) the ZZ metric, the cumulative distance between all 14 landmarks consecutively.

3.2. Comparing toothrow length metrics

Comparisons between the use of the traditional maximum toothrow length (MTL) and the cumulative toothrow length value (CTL) for quantifying the toothrow were made (Fig. 3A). Spacing Ratio crowding scores were calculated using both metrics, and the relationship between variables was tested using linear least-squared regressions and visualised with a scatterplot (Fig. 3B).

3.3. Comparing spacing and Rotation Ratios

To test which measurement best described crowding in the mandible, values produced using the Spacing Ratio (SR) and the Rotation Ratio (RR) were compared. The relationship between the two ratios was tested using linear least-squared regressions and visualised with a scatterplot (Fig. 4.

3.4. Examining crowding within groups

We tested the frequency of mandibular tooth crowding between groups with an analysis of variance (ANOVA) test, and visualised results with boxplots. A value of SR > 1.0 represents 'crowded' specimens, where the sum of the length of all teeth is greater than the length of the total toothrow (i.e. the specimen has no spaces or gaps in its toothrow). The prevalence of crowding was measured as the percentage of crowded specimens per group. A pairwise t-test

(with a Bonferroni correction for multiple comparisons) was used to examine variation between wild and domestic groups, including the subgroups of wolf-like breeds and Pariah dogs.

Differences between groups were examined using Canonical Variate Analysis (CVA), paired with leave-one-out cross validation and based on balanced groups to avoid bias linked with heterogeneous sample size (Evin et al., 2013). The leave-one-out cross validation method iteratively treats each specimen as an unknown and attempts to identify it. Thus, the percentage of correctly re-identified specimens provides an approximation for the ability to identify an unknown specimen to a given group, and is reported as a percentage of specimens correctly classified (cross-validation percentage; CVP). These analyses were applied to the complete dataset, as well as the subsets of the wolf-like breeds and Pariah dogs. CVA was conducted using both the Spacing Ratio and the Rotation Ratio measurements, to examine which (if any) method provides better discrimination between groups.

The variation within SR crowding scores due to differences in size was examined with a regression between the SR score and the toothrow centroid size. The centroid size (CS) was obtained after the Generalised Procrustes Superimposition (GPS) of the landmark coordinates. CS is calculated as the square root of the sum of the squared distances between each landmark and the centroid (mean of the landmarks), and is used as a univariate summary of overall size (Rohlf and Slice, 1990; Bookstein, 1991). In addition to the CS analysis, differences in toothrow shapes (calculated from new
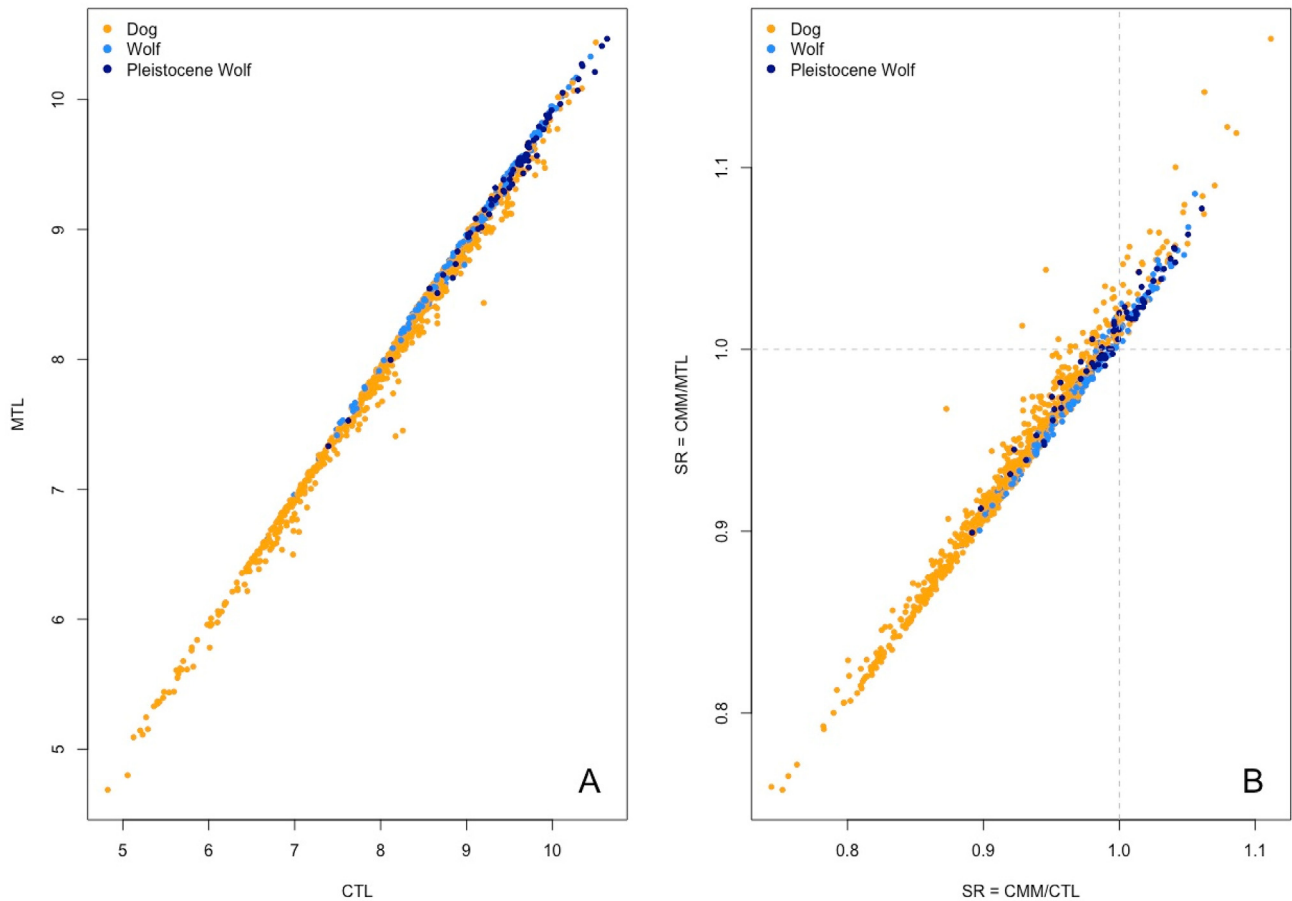
Fig. 3. A: Relationships between cumulative toothrow length (CTL) and maximum toothrow length (MTL). B: the SR crowding value calculated using each toothrow measurement, with dashed lines indicating a SR crowding value of 1.0.

coordinates obtained after the GPS) between groups were examined using a principal component analysis and discriminant analyses to determine if the shape data could provide a more detailed measure of biological variation due to tooth crowding (Adams et al., 2004; Rohlf and Marcus, 1993; Bookstein, 1996). The resulting principal component scores were individually compared with SR crowding scores to determine the correlation between crowding scores and toothrow morphology.

Since both wolves and domestic animals are known to exhibit sexual dimorphism (Crockford, 1997; Jolicoeur, 1975), the influence of sex on tooth crowding was examined with a two-way ANOVA to test whether sexual dimorphism influences the expression of tooth crowding among dog and wolves.

To test the influence of geography on crowding scores, the latitudinal position of each specimen was compared to the specimens SR crowding scores using a linear regression. Procrustean Randomisation test (PROtest) approach (Peres-Neto and Jackson, 2001) with 10,000 permutations was applied to measure the association between the geographic distance matrix (in km) and a similarity matrix (distance between individual SR crowding scores). This tests the hypothesis that specimens similar in SR crowding scores are also geographically close and, likewise, that specimens with different SR crowding scores are geographically distant. This was carried out on the global dataset as well as for each continent separately (North America and Eurasia), and p values were adjusted for multiple comparisons using the Bonferroni method.

All calculations of the previous measurements and analyses were executed in R (using the interface RStudio v0.99.467) (R Core Team, 2015). The author written R code for calculating the crowding metrics is available in the Supplementary information (SI 1).

4. Results

4.1. Measuring the toothrow

As expected, the MTL and CTL toothrow measurements are highly correlated (adjusted $R^2 = 0.995$, $p < 2.2e-16$, Fig. 3A). The Spacing Ratio crowding scores calculated with the two different toothrow length calculations (CTL and MTL) are also highly correlated ($p < 2.2e-16$), with an adjusted R^2 value of 0.8093 (Fig. 3B). Crowding scores calculated with the CTL measurement indicated that 10.38% of all specimens were crowded, whereas 18.60% of all specimens using the maximum toothrow length (MTL) measurement were considered crowded (Fig. 2B). A visual assessment of the outliers from this analysis revealed that the MTL measurement overestimated instances of crowding by failing to accurately describe specimens with less crowded teeth in more curved mandibles because it cannot account for the additional length introduced by the curvature. Thus, we used the cumulative toothrow length (CTL) for all further analysis as a more accurate representation of toothrow length.

4.2. Describing crowding — spacing ratio vs. Rotation Ratio

Crowding scores calculated using the Spacing Ratio and the Rotation Ratio were highly correlated (adjusted $R^2 = 0.8914$,
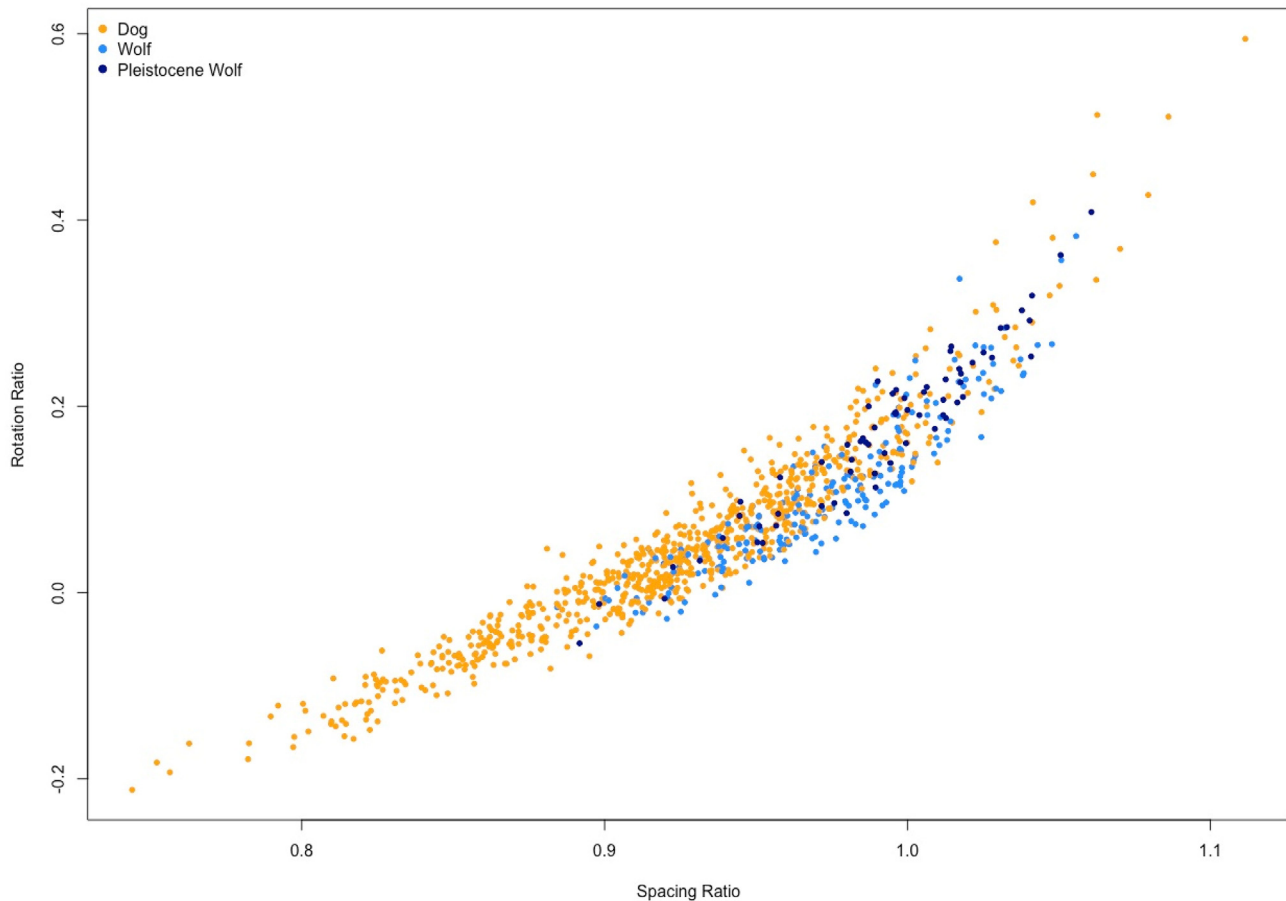
Fig. 4. Scatter plot of Spacing Ratio versus Rotation ratio values; showing the rapid increase of Rotation Ratio values relative to the Spacing Ratio scores in crowded specimens.

$p < 2.2e\text{-}16$). A scatter plot of the results of the Rotation Ratio against those of the Spacing Ratio shows how the Rotation Ration scores accentuated specimens with more crowding and attenuated the scores of specimens with a Spacing Ratio value less than 1 (Fig. 4).

4.3. Crowding in wolves and domestic dogs

The ANOVA test revealed significant overall differences between SR crowding scores in the modern dog, modern wolf and Pleistocene wolf groups ($p < 2.2e\text{-}16$). SR crowding scores indicate that 6.13% of modern dogs exhibited crowded mandibles (SR range: 0.74—1.11), while modern and Pleistocene wolves showed 18.04% (SR range: 0.84—1.07) and 36.36% (SR range: 0.89—1.06) crowding respectively (Fig. 5). For the subgroups of modern dogs, 9.57% of wolf-like specimens exhibited crowding (SR range: 0.74—1.08), as did 7.69% of Pariah dogs (SR range: 0.84—1.11). Pairwise *t*-tests on all five groups showed that the overall domestic dog group — as well as the wolf-like and Pariah subgroups — were significantly different from modern and Pleistocene wolves (at the $p < 0.05$ threshold) in crowding, but did not differ significantly from each other.

The CVA of SR crowding score, using balanced groups for 1000 iterations, had a mean CVP of 53.17% (Fig. 5). The separate comparison of the wolves (modern and Pleistocene) with wolf-like breeds and the Pariah dogs succeed to identify correctly 52.74% of specimens, and 51.13% respectively. These low CVP values indicate that SR scores should not be relied upon to separate wild and domestic groups.

Results of the CVA for the Rotation Ratio between all dogs, modern wolves, and Pleistocene wolves had a mean CVP of 51.14%

(on balanced groups for 1000 iterations) (SI Fig. 3). Separate comparisons of wolf-like breeds and Pariah dogs with the wolf groups, successfully identified 50.98% of the wolf-like specimens, and 51.50% of the Pariah dogs. The results of the CVA show that discrimination between groups is almost equal using the Spacing Ratio and Rotation Ratio calculations, suggesting that incorporating rotational changes into the crowding analysis does not influence the ability to identify between groups.

4.4. Shape analysis

Results of the CVA on shape coordinates indicate a mean CVP of 88.25% (on balanced groups for 1000 iterations) (SI Fig. 4). PC 1 is strongly correlated with SR scores (adjusted $R^2 = 0.6561$, $p > 0.001$). However, dogs and wolves overlapped extensively on PC 1. Shape changes visualised along PC 1 indicates that the variation includes the curvature of the toothrow, as well as the relative proportion of individual teeth (SI Fig. 5). Variation represented by PC 2 is predominately influenced by the curvature of the toothrow, and indicates greater differences between dogs and wolves (SI Fig. 6). This suggests that the more powerful discrimination between groups using geometric morphometric methods is due to the ability of these methods to account for more of the shape variation present in the toothrow, and that crowding represents only a small part of that overall variation.

4.5. Influence of size and sex

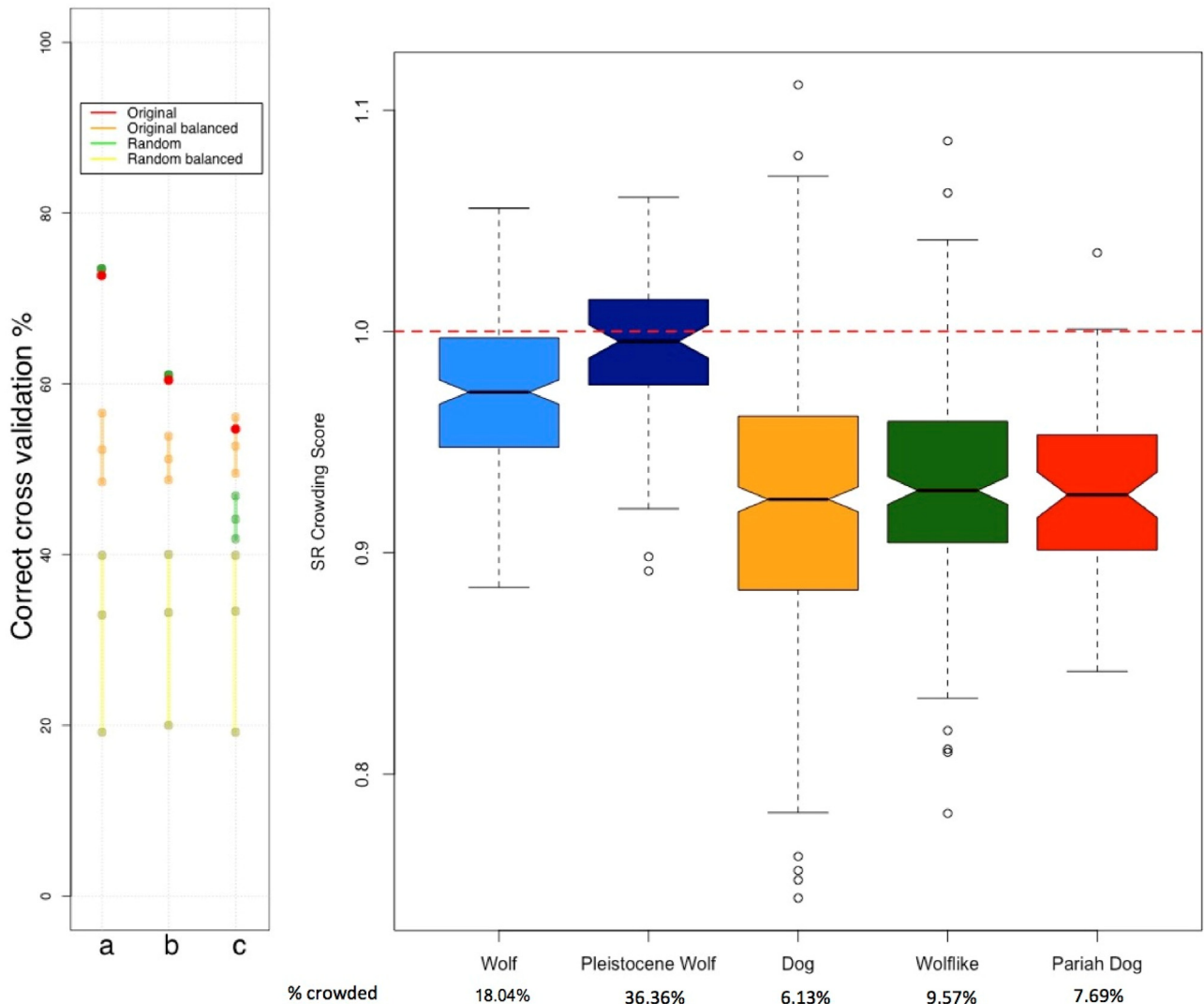The covariation between SR value and toothrow centroid size for

Fig. 5. (Left) Results of the CVA for identification between Modern Wolves, Pleistocene Wolves and (a) all modern Dogs, (b) Pariah Dogs, and (c) Wolf-like Breeds. Distributions are from 1000 iterations of each analysis. Red: CVP calculated on unbalanced (initial) samples, orange: on balanced samples, green: on random unbalanced (initial) samples, yellow: random and balanced samples. (Right) Boxplot showing SR values for modern wolves, Pleistocene wolves, all modern dogs, and subgroups of "Wolf-like" breeds and pariah dogs. Dashed line indicates SR values of 1.0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

all specimens indicated a weak correlation between CS and tooth crowding (adjusted $R^2 = 0.027$, $p < 0.001$). Each group was examined individually, and similarly weak correlations were reported for modern wolf (adjusted $R^2 = 0$, $p = 0.819$), Pleistocene wolf (adjusted $R^2 = 0.003$, $p = 0.283$), and dog groups (adjusted $R^2 = 0.084$, $p < 0.001$), while a slightly stronger correlation was found in the modern dog subgroups of Pariah dogs (adjusted $R^2 = 0.289$, $p < 0.001$) and wolf-like breeds ($r = 0.415$, $p < 0.001$).

The results of a two-way ANOVA showed no significant difference in tooth crowding between males and females for both dog (F = 0.267; $p = 0.6055$) and modern wolf (F = 0.0357; $p = 0.85504$) groups.

4.6. Geographic variation in modern wolves

Tooth crowding was found to have a significant ($p < 0.001$), but weak positive correlation (adjusted $R^2 = 0.054$) with latitudinal change. The results of the global Procrustes superimposition were non-significant over the whole geographic range of the dataset ($p = 0.81$) and within each continent (Eurasia $p = 0.67$; North America

$p = 0.32$). This suggests that the influence of geography on the variation of tooth crowing does not need to be considered.

5. Discussion

5.1. Methods for recording tooth crowding

Our results demonstrate that landmark based methods are an effective tool for measuring tooth crowding in dogs and wolves, and that the cumulative toothrow measurement (CTL) provides a more comprehensive description of the canid mandibular toothrow than the traditional linear measurement (MTL). The MTL fails to accurately describe specimens with curved mandibles and, as a result, inflates the crowding value in specimens with more curved mandibles. Previous analyses of tooth crowding which used similar linear measurements to define the mandibular toothrow have likely over-estimated the amount of crowding in their samples.

There was no previous consensus on what index value constitutes a threshold to classify a specimen as crowded. For example, Benecke (1994) analysis of wolves from the Gravettian site of

Předmostí reports the presence of crowding in 4% of recent wolves, with an average crowding index of 99.4%. However, the average crowding index for Předmostí specimens classified as "uncrowded" in the same study was also very high (91.2%). Based on the results presented here, we propose that a Spacing Ratio value of 1.0 should be used as a simple threshold for describing overall crowding. This value represents specimens where the length of the toothrow and the cumulative length of the mandibular teeth are equal. Since values lower than 1.0 indicate relatively less densely packed teeth than those values closer to or greater than 1.0, specimens can be examined along this spectrum to quantitatively study the spacing of mandibular teeth and the relative proportion of crowding observed. This definition provides a systematic benchmark for interpreting crowding (or lack thereof) among and between wild and domestic populations, since specimens with a spacing ratio value of ≥ 1.0 can be confidently considered as 'crowded.'

5.2. Domestication and tooth crowding

Our results provide the first comprehensive evidence that mandibular tooth crowding occurs in ancient and modern wolf populations as well as in modern domestic dogs. The highest proportion of crowding is present in the wild wolf groups, as Pleistocene wolves from Alaska exhibit almost double the percentage of crowding compared to modern wolves. While dogs exhibited the lowest overall proportion of crowding, they had the greatest variation in crowding scores, as well as the single highest overall SR crowding value. Prior to this study, crowding in wolf populations was considered to be restricted to captive wolves and those under human management (Degerbøl, 1961; Lawrence, 1967; van Wijngaarden-Bakker, 1974), or had been variously attributed to the effects of hybridisation and/or introgression between dogs and wolves (Dimitrijević and Vuković, 2012; Koler-Matznick, 2002). While we cannot exclude the possibility of wolf-dog hybridisation as an explanation for crowding (hybridisation being well documented, see Vilà and Wayne, 1999; Moura et al., 2014; Randi, 2011; Verlag and Service, 2002; Crockford, 2000), the large size of our dataset, along with the inclusion of Pleistocene wolves, minimises any likely influences. Interestingly, the highest proportion of crowded specimens was observed in the Late Pleistocene wolves from Alaska. These specimens mostly pre-date the arrival of humans in North America (Goebel et al., 2008; although see Bourgeon et al., 2017) adding support to our observation that a degree of mandibular tooth crowding should be expected in the range of natural ecomorphs represented within Canis lupus.

While no crowding index value has been (nor should be) definitively shown to determine wild/domestic status, and previous studies have similarly recorded this large overlap in crowding scores between wild and domestic specimen (e.g. Degerbøl, 1961; Benecke, 1994), trends observed in the crowding of wild and domestic specimens have frequently been generalised to support the identification of early domestic dog (see Introduction). Our results indicate a seemingly higher frequency of crowding in both modern and ancient wolves, strongly suggesting that the reasons for its occurrence is more complex than a simple wild-domestic dichotomy.

Although we found very little geographic signal from the distribution of SR crowding scores in modern wolves, there are several biological trends and rules that link size and morphology with latitude (e.g. Bergmann's rule, Meiri et al., 2003), and it is possible that these may influence tooth crowding in wolves. Such investigation is beyond the remit of this paper, however future studies should consider this, and focus on examining the functional and morphological development of crowding in extant populations to better understand the underlying drivers of tooth crowding.

The range of domestic dog SR crowding values encompasses all values reported by our modern and Pleistocene wolves (Fig. 5). Given this extensive overlap, we conclude that there is no crowding score or range of scores that can be confidently used to identify domestic dogs. Similarly, though the discriminate analysis on shape coordinates provided good discrimination between wild and domestic groups (SI Fig. 4), this analysis incorporated morphological characteristics beyond crowding, including not only the curvature of the toothrow but also position and relative size of the teeth. Thus, tooth crowding alone is a poor criterion for the identification of domestic dogs.

5.3. Wolf-like breeds, pariah dogs and domestic morphology

Reference to modern wolf-like breeds, also referred to as "northern" breeds, and "archaic" breeds, is common in research exploring the origins of dog domestication (e.g. Parker et al., 2004; Leonard et al., 2002; Germonpré et al., 2009; Larson et al., 2012; Skoglund et al., 2015; see Table 1). These breeds are often considered to be relatively "unmodified", meaning that they retain characteristics most similar to wolves (Lawrence, 1967:57). Similarly, Pariah dogs in particular are thought to reflect an average morphology less influenced by selective breeding (Drake and Klingenberg, 2010) and have retained (or perhaps regained) more of the "natural" morphological characteristics supposedly exhibited by early domesticates (Coppinger and Coppinger, 2001). Our results demonstrate that both the wolf-like and the Pariah groups were distinct from modern and Pleistocene wolf populations, but not from each other, nor from the modern dog group.

These results further support the conclusion that wild populations exhibit more crowding than domestic ones, and that there is no evidence to suggest that it is possible to discriminate between wild and domestic groups based on crowding alone — even in those breeds thought to best reflect proto-domestic morphologies.

5.4. Anatomy of the toothrow

Alignment and rotational changes of the teeth due to crowding have been analysed in both the canid maxilla (Degerbøl, 1961; Brothwell, 1991) and the mandible (Dimitrijević and Vuković, 2012), and the recording and quantification of this rotation (along with the influence of different measurements of the toothrow on crowding scores) is worth considering in the context of the overall anatomy. Maxillary tooth crowding often takes on a different form to that exhibited in the mandible, and crowding in the maxilla usually results in the rotation of the third premolar and the formation of a step between the fourth and second premolar. This results in a maxillary toothrow that appears more curved. As a result, if a cumulative method of quantifying the maxillary toothrow (similar to CTL) was used in such an instance, the curvature (induced by the crowding) would be minimised, and consequently the results would likely underestimate the degree of tooth crowding in the maxilla.

These findings illustrate the importance of examining (and attempting to understand) the nature of tooth crowding on the development and anatomy of the specific element and taxa under study. Since the modularity of the canid jaw and associated tooth position is important for biological and ecological studies of evolution and adaptation as an indicator of diet, food acquisition strategies and environmental specialisation (Meloro et al., 2008; Crusafont-Pairó and Truyols-Santonja, 1956) further analyses could help to explain the occurrences of tooth crowding observed in wild canids and other carnivore species.

6. Conclusions

This study represents the first comprehensive examination of mandibular tooth crowding in a large geographical and temporal dataset of dogs and wolves. Our results show firstly that systematic quantification of tooth position in the canid mandible is achievable using landmark-based metric analysis, and that these metrics can be used to quantitatively explore instances of tooth crowding in canids. In addition, the Spacing Ratio measurement provides a robust quantification of mandibular tooth crowding, and measuring the toothrow by accounting for curvature of the hemimandible is crucial to avoid overestimating crowding.

While there were significant differences in crowding between wolves and domestic dogs, the results of discriminant analyses provided low cross-validation values, demonstrating that wolves and domestic dogs cannot be confidently separated based on tooth crowding alone, contrary to the geometric morphometric approach which incorporates more of the morphology of the toothrow, and appeared a promising tool for discriminating between groups. The fact that significantly high instances of mandibular tooth crowding are present in modern and Late Pleistocene wild wolf populations relative to dogs contradicts the commonly reported assumption that tooth crowding can be used both to discriminate dogs from wolves, and as a marker of the early domestication process. These results suggest that status determinations of dogs and other domesticates based upon tooth crowding alone should be treated with caution, and that other methods including landmark-based geometric morphometrics should be explored.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jas.2017.06.014.

References

Adams, D., Rohlf, F.J., Slice, D., 2004. Geometric morphometrics: ten years of progress following the "revolution". Ital. J. Zool. 71, 5–16.

Benecke, N., 1987. Studies on early dog remains from northern Europe. J. Archaeol. Sci. 14, 31–50.

Benecke, N., 1994. Archäozoologische Studien zur Entwicklung der Haustierhaltung. Akademie Verlag, Berlin.

Bookstein, F.L., 1991. Morphometric Tools for Landmark Data. Geometry and Biology. Cambridge University Press, New York.

Bookstein, F.L., 1996. Biometrics, biomathematics and the morphometric synthesis. Bull. Math. Biol. 58, 313–365.

Boudadi-Maligne, M., Escarguel, G., 2014. A biometric re-evaluation of recent claims for Early Upper Palaeolithic wolf domestication in Eurasia. J. Archaeol. Sci. 45, 80–89.

Bourgeon, L., Burke, A., Higham, T., 2017. Earliest human presence in North America dated to the last glacial maximum: new radiocarbon dates from bluefish caves, Canada. Plos one 12 (1), e0169486. http://dx.doi.org/10.1371/journal.pone.0169486.

Brothwell, D., 1991. Malocclusion and methodology: the problem and relevance of recording dental malalignment in archaeology. Int. J. Osteoarchaeol 1, 27–37.

Clark, K.M., 1996. Neolithic Dogs : a reappraisal based on evidence from the remains of a large canid deposited in a ritual feature. Int. J. Osteoarchaeol 6, 211–219.

Clutton-Brock, J., 1963. The origins of the dog. In: Brothwell, D., Higgs, E. (Eds.), Science in Archaeology: a Comprehensive Survey of Progress and Research, second ed. Praeger, New York, pp. 303–309.

Clutton-Brock, J., 1999. The processes of domestication. Mammal. Rev. 22 (2), 79–85.

Coppinger, R., Coppinger, L., 2001. Dogs: a New Understanding of Canine Origins, Behaviour and Evolution. Scribner, New York.

Crockford, S.J., 1997. Osteometry of Makah and Coast Salish Dogs. Archaeology Press. Simon Fraser University, Burnaby, B.C.

Crockford, S.J., 2000. A commentary on dog evolution: regional variation, breed development and hybridisation with wolves. In: Crockford, S.J. (Ed.), Dogs Through Time: an Archaeological Perspective. Archaeopress, Oxford, pp. 295–312.

Crockford, S.J., Kuzmin, Y.V., 2012. Comments on Germonpre et al., Journal of Archaeological Science 36, 2009 "Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes", and Germonpre, Lazkickova-Galetova, and Germonpre, L. J. Archaeol. Sci. 39 (8), 2797–2801.

Crusafont-Pairó, M., Truyols-Santonja, J., 1956. A biometric study of evolution of fissiped carnivores. Evolution 10, 314–332.

Darwin, C., 1868. The Variation of Plants and Animals under Domestication, vol. 1. Appleton, New York.

Davis, S.J.M., Valla, F.F., 1978. Evidence for domestication of the dog 12,000 years ago in Natufian of Isreal. Nature 276, 609–610.

Degerbøl, M., 1961. On a find of a Preboreal domestic dog (Canis familiaris L.) from Star Carr, Yorkshire, with remarks on other mesolithic dogs. Proc. Prehist. Soc. 21, 35–55.

Dimitrijević, V., Vuković, S., 2012. Was the dog locally domesticated in the Danube Gorges? Morphometric study of dog cranial remains from four mesolithic-early neolithic archaeological sites by comparison with contemporary wolves. Int. J. Osteoarchaeol 25 (1), 1–30. http://dx.doi.org/10.1002/oa.2260.

Drake, A.G., 2011. Dispelling Dog Dogma: an investigation of heterochrony in using 3D geometric morphometric analysis of skull shape. Evol. Dev. 13 (2), 204–213.

Drake, A.G., Klingenberg, C.P., 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. Am. Nat. 175 (3), 289–301.

Evin, A., Cucchi, T., Cardini, A., Vidarsdottir, U.S., Larson, G., Dobney, K., 2013. The long and winding road: identifying pig domestication through molar size and shape. J. Archaeol. Sci. 40 (1), 735–743.

Germonpre, M., Sablin, M.V., Stevens, R.E., Hedges, R.E., Hofreiter, M., Stiller, M., Despres, V.R., 2009. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. J. Archaeol. Sci. 36 (2), 473–490.

Germonpre, M., Laznickova-Galetova, M., Sablin, M.V., 2012. Palaeolithic dog skulls, at the gravettian predmostí site, the Czech Republic. J. Archaeol. Sci. 39 (1), 184–202.

Germonpré, M., Sablin, M.V., Despres, V., Hofreiter, M., Laznickova-Galetova, M., Stevens, R.E., Stiller, M., 2013. Palaeolithic dogs and the early domestication of the wolf: a reply to the comments of Crockford and Kuzmin (2012). J. Archaeol. Sci. 40 (1), 786–792.

Germonpré, M., Sablin, M.V., Laznickova-Galetova, M., Despres, V., Stevens, R.E., Stiller, M., Hofreiter, M., 2015a. Palaeolithic dogs and pleistocene wolves revisited : a reply to Morey (2014). J. Archaeol. Sci. 54, 210–216.

Germonpré, M., Laznickova-Galetova, M., Losey, R., Raikkonen, J., Sablin, M.V., 2015b. Large canids at the Gravettian Předmostí site, the Czech Republic: the mandible. Quat. Int. 359–360, 1–19.

Goebel, T., Waters, M.R., O'Rourke, D.H., 2008. The late Pleistocene dispersal of modern humans in the americas. Science 319 (5869), 1497–1502.

Jolicoeur, P., 1975. Sexual dimorphism and geographic distance as factors of skull variation in the wolf (Canis lupus). In: Fox, M.W. (Ed.), The Wild Canids: Their Systematics, Behavioural Ecology and Evolution. Van Nostrand Reinhold Co., New York, pp. 54–61.

Koler-Matznick, J., 2002. The origin of the dog revisited. Anthrozoos 15, 98–118.

Krause-Kyora, B., Makarewicz, C., Evin, A., Flink, L.G., Dobney, K., Larson, G., Hartz, S., Schreiber, S., von Carnap-Bornheim, C., von Wurmb-Schwark, N., Nebel, A., 2013. Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. Nat. Commun. 4, 2348.

Lapham, H.A., 2010. A baumer phase dog burial from the kincaid site in southern Illinois. Ill. Archaeol. Surv. Ill. Archaeol. 22 (2), 437–463.

Larson, G., Karlsson, E.K., Perri, A., Webster, M.T., Ho, S.Y., Peters, J., Stahl, P.W., Piper, P.J., Lingaas, F., Fredholm, M., Comstock, K.E., Modiano, J.F., Schelling, C., Agoulnik, A.I., Leegwater, P.A., Dobney, K., Vigne, J.D., Vila, C., Andersson, L., Lindblad-Toh, K., 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. PNAS 109 (23), 8878–8883.

Lawrence, B., 1967. Early domestic dogs. Z. für Säugetierkd. 32, 44–59.

Leonard, J.A., Wayne, R., Wheeler, J., Valdez, R., Guillen, S., Vila, C., 2002. Ancient DNA evidence for old world origin of new world dogs. Science 298, 1613–1616.

Leonard, J.A., Vila, C., Fox-Dobbs, K., Koch, P.L., Wayne, R.K., Van Valkenburgh, B., 2007. Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. Curr. Biol. 17 (13), 1146–1150.

Meiri, S., Dayan, T., Aviv, T., 2003. On the validity of Bergmann's rule. J. Biogeogr. 30, 331–351.

Meloro, C., et al., 2008. The shape of the mandibular corpus in large fissiped carnivores: allometry, function and phylogeny. Zool. J. Linn. Soc. 154 (4), 832–845.

Moray, D.F., 1994. The early evolution of the domestic dog. Am. Sci. 82, 336–347.

Morey, D.F., 1992. Size, shape and development in the evolution of the domestic

dog. J. Archaeol. Sci. 19 (2), 181—204.

Morey, D.F., 2014. In Search of paleolithic dogs: a quest with mixed results. J. Archaeol. Sci. 52, 300—307.

Moura, A.E., Tsingarska, E., Dabrowski, M.J., Czarnomska, S.D., Jedrzejewska, B., Pilot, M., 2014. Unregulated hunting and genetic recovery from a severe population decline: the cautionary case of Bulgarian wolves. Conserv. Genet. 15, 405—417.

Musil, R., 2000. Evidence for the domestication of wolves in central European magdalenian sites. In: Crockford, S.J. (Ed.), Dogs through Time: an Archaeological Perspective. Archaeopress, Oxford, pp. 295—312.

Napierala, H., Uerpmann, H.-P., 2012. A 'new' paleolithic dog from central Europe. Int. J. Osteoarchaeol 22, 127—137.

Ovodov, N.D., Crockford, S.J., Kuzmin, Y.V., Higham, T.F., Hodgins, G.W., van der Plicht, J., 2011. A 33,000-year-old incipient dog from the Altai Mountains of Siberia: evidence of the earliest domestication disrupted by the last glacial maximum. PLoS One 6 (7), 4—10.

Parker, H.G., Kim, L.V., Sutter, N.B., 2004. Genetic structure of the purbred domestic dog. Science 304, 1160—1164.

Peres-Neto, P., Jackson, D., 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia 129, 169—178.

Perri, A., 2016. A wolf in dog's clothing: initial dog domestication and Pleistocene wolf variation. J. Archaeol. Sci. 68, 1—4.

R Core Team, 2015. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project. org/.

Randi, E., 2011. Genetics and conservation of wolves Canis lupus in Europe. Mamm.

Rev. 41, 99—111.

Rohlf, F.J., 2013. tpsDig. In: Ecology and Evolution. State University of New York at Stony Brook, Stony Brook, NY.

Rohlf, J.F., Marcus, L.F., 1993. A revolution morphometrics. Trends Ecol. Evol. 8, 129—132.

Rohlf, F.J., Slice, D.E., 1990. Extensions of the procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39, 40—59.

Sablin, M.V., Khlopachev, G.A., 2002. The earliest ice age dogs: evidence from eliseevichi 1. Curr. Anthropol. 43 (5), 795—799.

Skoglund, P., Ersmark, E., Palkopoulou, E., Love, D., 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr. Biol. 25 (11), 1515—1519.

van Wijngaarden-Bakker, L.H., 1974. The animal re- mains from the Beaker settlement at Newgrange, Co. Meath: first report. Proc. R. Ir. Acad. 74, 313—383.

Verlag, F., Service, F., 2002. Hybridisation between wolves and dogs in Latvia as documented using mitochondrial and microsatellite DNA markers. Mamm. Biol. 67 (2), 79—90. https://doi.org/10.1078/1616-5047-00012.

Vilà, C., Wayne, R.K., 1999. Hybridization between wolves and dogs. Conserv. Biol. 13, 195—198.

Walker, D.N., Frison, G.C., 1982. Studies on amerindian dogs, 3: prehistoric wolf/dog hybrids from the northwestern plains. J. Archaeol. Sci. 9, 125—172.

Wayne, R., 1986. Cranial Morphology of wild and domestic canids. The influence of development on morphological change. Evolution 40 (2), 243—261.

Zeder, M.A., 2012. Pathways to animal domestication. In: Gepts, P., Famula, T., Bettinger, R., Brush, S., Damania, A., McGuire, P., Qualset, C. (Eds.), Biodiversity in Agriculture: Domestication, Evolution, and Sustainability. Cambridge University Press, Cambridge, pp. 227—259.