The Anatomical Record WILEY

Morphological evidence for early dog domestication in the European Pleistocene: New evidence from a randomization approach to group differences

Patrik Galeta[1] | Martina Lázničková-Galetová[1,2] | Mikhail Sablin[3] | Mietje Germonpré[4]

[1]Department of Anthropology, University of West Bohemia, Pilsen, Czech Republic

[2]The Moravian Museum, Brno, Czech Republic

[3]Zoological Institute, Russian Academy of Sciences, Saint-Petersburg, Russia

[4]Operational Direction "Earth and History of Life", Royal Belgian Institute of Natural Sciences, Brussels, Belgium

Correspondence
Patrik Galeta, Department of Anthropology, University of West Bohemia, Sedláčkova 15, 306 14 Pilsen, Czech Republic.
Email: galeta@ksa.zcu.cz

Abstract

The antiquity of the wolf/dog domestication has been recently pushed back in time from the Late Upper Paleolithic ($\sim$14,000 years ago) to the Early Upper Paleolithic (EUP; $\sim$36,000 years ago). Some authors questioned this early dog domestication claiming that the putative (EUP) Paleolithic dogs fall within the morphological range of recent wolves. In this study, we reanalyzed a data set of large canid skulls using unbalanced- and balanced-randomized discriminant analyses to assess whether the putative Paleolithic dogs are morphologically unique or whether they represent a subsample of the wolf morpho-population. We evaluated morphological differences between 96 specimens of the 4 a priori reference groups (8 putative Paleolithic dogs, 41 recent northern dogs, 7 Pleistocene wolves, and 40 recent northern wolves) using discriminant analysis based on 5 ln-transformed raw and allometrically size-adjusted cranial measurements. Putative Paleolithic dogs are classified with high accuracies (87.5 and 100.0%, cross-validated) and randomization experiment suggests that these classification rates cannot be exclusively explained by the small and uneven sample sizes of reference groups. It indicates that putative Upper Paleolithic dogs may represent a discrete canid group with morphological signs of domestication (a relatively shorter skull and wider palate and braincase) that distinguish them from sympatric Pleistocene wolves. The present results add evidence to the view that these specimens could represent incipient Paleolithic dogs that were involved in daily activities of European Upper Paleolithic forager groups.

KEYWORDS

discriminant analysis, dog, domestication, Europe, morphology, Pleistocene, randomization, wolf, zoology

1 | INTRODUCTION

The issue of dog domestication in the European Late Pleistocene has been the subject of heated debate during the past several years (Botigué et al., 2017; Crockford & Kuzmin, 2012; Drake, Coquerelle, & Colombeau, 2015; Frantz et al., 2016; Germonpré et al., 2009, 2013; Germonpré, Lázničková-Galetová, Losey, Räikkönen, & Sablin, 2015; Germonpré, Laznickova-Galetova, & Sablin, 2012; Germonpré, Sablin, et al., 2015; Grimm, 2015;

Morell, 2015; Morey, 2014). While it is accepted that the sole ancestor of the dog is the wolf, disagreement prevails about the dating of the domestication process (Botigué et al., 2017; Frantz et al., 2016; Germonpré et al., 2009, 2012, 2013; Germonpré, Sablin, et al., 2015; Perri, 2016). Dog domestication is safely dated back to the Late Upper Paleolithic (LUP, ca. 14,000 years ago) (e.g., Street, Napierala, & Janssens, 2015; Thalmann et al., 2013). Recently, however, morphological studies (Germonpré et al., 2009) suggested that the earliest evidence of wolf domestication is much older and that this process could have started in the Early Upper Paleolithic (EUP, ca. 36,000 years ago).

The morphological evidence for the early domestication of the dog was based on the discriminant analyses (DA) of cranium (Germonpré et al., 2009, 2013) and mandible measurements (Germonpré, Lázničková-Galetová, et al., 2015). In the article from 2013, for example, Germonpré et al. combined seven cranial measurements of four large canid reference groups (recent archaic dogs, putative Paleolithic dogs, recent and Pleistocene wolves) and showed that analysis computed from both raw and size-adjusted data provided good discrimination among groups with a high rate of correct classification (97 and 92%, non-cross-validated). As a result, Germonpré et al. argued for the existence of two distinct morpho-populations (a group of individuals distinguishable from other populations of the species by morphometric characteristics; previously called morphotypes) of EUP large canids: Paleolithic dogs (Figure 1) and Pleistocene wolves. Compared to the Pleistocene wolf morpho-population, members of the Paleolithic dog morpho-population have a shorter skull, a shorter and relatively broader snout, a relatively wider braincase, and a shorter lower jaw (Germonpré et al., 2009, 2012, 2017; Germonpré, Lázničková-Galetová, et al., 2015).

The existence of a Paleolithic dog morpho-population has been questioned from different perspectives. Several

FIGURE 1 Lateral view of the Paleolithic dog from the Goyet Cave, Belgium, Early Upper Paleolithic, ca. 35,500 years ago

authors asserted that morphological evidence for its domestic nature is ambiguous (Boudadi-Maligne & Escarguel, 2014; Crockford & Kuzmin, 2012; Napierala & Uerpmann, 2012). They argued that the overall skull size of the putative Paleolithic dogs is the same as the skull size of recent wolves and simultaneously is well larger than the size of recent dogs (Boudadi-Maligne & Escarguel, 2014; Morey, 2014). Drake et al. (2015) claimed that Paleolithic dogs lack typical morphological traits of their domesticated counterparts such as cranial flexion and concavity near the orbits. Janssens, Perri, Crombé, Van Dongen, and Lawler (2019) considered that the small body size of the Paleolithic dogs can be solely explained by the low sample size of this group. According to these authors, Paleolithic dogs do not represent a discrete morphological group and fall within the natural morphological variability of modern wolves.

Some authors argued (Morey, 2014) that the identification of the Paleolithic dog group was based on inappropriate methods of size adjustment, which led to misleading interpretation of the shape differences between both EUP morpho-populations. To reduce the effect of size, Germonpré et al. (2009); Germonpré, Lázničková-Galetová, et al. (2015) used an isometric size adjustment method (Jungers, Falsetti, & Wall, 1995), that is, considered that shape is preserved among specimens of different sizes. On the computational level, they divided each raw variable by the overall size of a specimen, which was expressed as a geometric means of all variables (Germonpré et al., 2012). By contrast, Morey (1992) considered that the shape of canids changes as a function of size and propose the use of nonisometric, that is, allometric size adjustment methods.

A further issue with distinguishing the Paleolithic dog morpho-population relates to the possible violation of discriminant analysis assumptions. As a rule of thumb, it is suggested that the sample size of the smallest reference group should be larger than the number of predictor variables (Kovarovic, Aiello, Cardini, & Lockwood, 2011; Tabachnick & Fidell, 2013) or even several times larger than the number of predictors (Ousley & Jantz, 2012). In Germonpré et al. (2009); however, the Paleolithic dogs, the smallest reference group, had only five specimens, which is less than seven cranial measurements used as predictors and substantially lower than the sample size of other reference groups. An impact of small sample size and sample size heterogeneity among reference groups on the accuracy of DA results has not been fully explored in the analyses of EUP large canids.

Small sample size, unbalanced composition of reference groups, and ambiguity in the interpretation of morphological differences between the Paleolithic dog and Pleistocene wolf morpho-populations make the use of

randomization methods desirable to explore issues with the early dog domestication (see other applications of the randomization to DA in zoology and anthropology in Cardini & Elton, 2011; Evin et al., 2013; Galeta, Bruzek, & Láznicková-Galetová, 2014; Kovarovic et al., 2011). The idea of the randomization experiment is to generate many randomized data sets with the same (unbalanced) sample composition as in the original data set but with group differences eliminated by random affiliation (see below) and then to assess whether DA based on the original data set performs better than the baseline estimated by randomization.

This study tries to push the debate on morphological changes in the canid domestication during EUP forward by addressing much of the criticism of the methods used in earlier investigations (Germonpré et al., 2009, 2012). To examine morphological differences among large canids, we employed between-group principal component analysis (bgPCA), cluster analysis, and discriminant analysis based on ln-transformed raw and allometrically size-adjusted craniometric variables of four canid groups (putative Paleolithic dogs, recent northern dogs, Pleistocene wolves, and recent [Old World] northern wolves). We adopted a randomization approach to the discriminant analysis to assess whether the putative Paleolithic dog group represents a distinct morpho-population of canids or whether they are in fact morphologically indiscernible from wolves. Specifically, we examined whether discriminant analyses showing differences between putative Paleolithic dogs and wolves (cf. Germonpré et al., 2009, 2012, 2013) can be explained by stochastic variation in DA results due to small and unbalanced reference group sizes.

2 | MATERIALS AND METHODS

2.1 | Large canid reference groups

The data set consisted of 96 skulls of recent and fossils large canids (Table 1). Following the previous works of Germonpré et al. (2009, 2012), the sample was divided into four a priori reference groups: putative Paleolithic dogs (PDs; $n = 8$), RNDs ($n = 41$), Pleistocene wolves (PWs; $n = 7$), and recent northern wolves (RNWs; $n = 40$). Note that Paleolithic dog and Pleistocene wolf groups were named for consistency with the previous literature (e.g., Germonpré et al., 2009), but their status is a matter of debate.

The putative Paleolithic dog morpho-population was composed of skulls discovered at several major Upper Paleolithic sites in three European regions dating from three subsequent time periods: (a) one Western

European specimen with an Aurignacian age (Goyet 2860, Belgium, calibrated age: ca. 35,500 years BP); (b) three Central European specimens dating from the Gravettian, these three were found at the Czech Předmostí site (Předmostí [−], 1060 and 1063, calibrated age: ca. 28,500 years BP); and (c) four Eastern European specimens, from the Russian Plain, dating from the Epigravettian (Mezhirich 4493 and Mezin 5490, Ukraine, estimated calibrated age: 18,000 years BP and Eliseevichi 447 and 23781, Russia, calibrated age: ca. 16,500 years BP). The Pleistocene wolf morpho-population is a more disparate ensemble consisting of seven specimens, most were discovered at Upper Paleolithic sites, a few at natural sites, in Belgium (Trou des Nutons), France (Maldidier), the Czech Republic (Předmostí), The Ukraine (Mezin), and Russia (Yakutia and Russian Plain) (Germonpré et al., 2017).

The RND group contained native dogs from Yakutia, Chukotka, Sakhalin Island, and Greenland, dating from the 19th and 20th centuries. These dogs, living in the arctic and subarctic, are, just as the Pleistocene large canids were, adapted to a cold environment (see also Germonpré, Sablin, et al., 2015). Importantly, the amount of recent admixture between the recent autochthonous northern dogs and modern European breeds was likely limited thanks to the geographic and cultural isolation of the former (Brown, Darwent, & Sacks, 2013; Larson et al., 2012; van Asch et al., 2013).

The RNW group was composed of Palaearctic wolves that lived in the 19th, 20th, or 21st century in Belgium, Sweden, and in several regions in Russia (Russian Plain, Yamal, Yakutia, Kamchatka, Far East). Several of these regions (Sweden, Yamal, Yakutia, Kamchatka) are at higher (>55N°) latitudes with a taiga or tundra biome and are, just like RNDs and the Pleistocene large canids of our sample, adapted to a cold environment. All fossil and prehistoric sites from which large canids were examined in this contribution are situated in the large geographical range of the RNW group. We did not include recent North American wolves in this reference group since the North American wolf lineage separated before the lineages of modern Eurasian wolves and domestic dogs diverged (Fan et al., 2016; Pilot et al., 2019) and because none of the studied sites is located in North America.

2.2 | Cranial measurements

Seven linear cranial and dental measurements were taken from each skull: TL (No. 1), total skull length; VL (No. 8), viscerocranium length; AL (No. 15), alveolar length of the tooth row P1–M2; P4CL (No. 18),

TABLE 1 Composition of the data set of large canids (*n* = 96)

Reference group	Period	Archeology	Age/AMS	Institute	References	*n*
Paleolithic dogs (PDs, *n* = 8)						
Goyet 2860 (Belgium)	LP	Aurignacian	31,890 + 240/ −220 BP 31,680 ± 250 BP	RBINS	Germonpré et al. (2009, 2012)	1
Předmostí (Czech Republic)	LP	Gravettian		MZM	Germonpré et al. (2012, 2013)	3
Mezin (Ukraine)	LP	Epigravettian		PM NASU	Benecke (1987), Germonpré et al. (2009), and Pidoplichko (1998)	1
Mezhirich (Ukraine)	LP	Epigravettian		PM NASU	Benecke (1987), Germonpré et al. (2009), and Pidoplichko (1998)	1
Eliseevichi (Russia)	LP	Epigravettian	13,905 ± 55 BP	MAE RAS, ZIN RAS	Germonpré et al. (2009, 2012) and Sablin and Khlopachev (2002)	2
Recent northern dogs (RNDs, *n* = 41)						
Sakhalin dog (Russia)	HOL		19th c.	SNM	Germonpré et al. (2009, 2012), Germonpré, Láznicková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	1
Siberian dogs (Russia)	HOL		19–20th c.	ZIN RAS	Germonpré et al. (2009, 2012), Germonpré, Láznicková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	19
Inuit dogs (Greenland)	HOL		19–20th c.	SNM	Germonpré, Láznicková-Galetová, et al. (2015) and Germonpré, Sablin, et al. (2015)	21
Pleistocene wolves (PWs, *n* = 7)						
Trou des Nutons (Belgium)	LP		21,810 ± 90 BP	RBINS	Germonpré et al. (2009)	1
Maldidier (France)	LP				Boudadi-Maligne and Escarguel (2014)	1
Předmostí (Czech Republic)	LP	Gravettian		MZM	Germonpré et al. (2012, 2013)	1
Mezin (Ukraine)	LP	Epigravettian		PM NASU	Germonpré et al. (2009)	2
Kostenki-17/II (Russia, plain)	LP	EU Paleolithic		ZIN RAS	Germonpré et al. (2012)	1
Anabar (Russia, Yakutia)	LP			ZIN RAS	Germonpré et al. (2009)	1
Recent northern wolves (RNWs; *n* = 40)						
Belgium	HOL		19th c.	RBINS	Germonpré et al. (2009, 2012), Germonpré, Láznicková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	1
Sweden	HOL		21st c.	NRM	Germonpré, Láznicková-Galetová, et al. (2015) and Germonpré, Sablin, et al. (2015)	3
Russia, Russian plain	HOL		19–20th c.	ZIN RAS	Germonpré et al. (2009, 2012), Germonpré, Láznicková-Galetová, et al., (2015), and Germonpré, Sablin, et al., (2015)	16

TABLE 1 (Continued)

Reference group	Period	Archeology	Age/AMS	Institute	References	*n*
Russia, Yamal	HOL		20th c.	ZIN RAS	Germonpré et al. (2009, 2012), Germonpré, Lázničková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	6
Russia, Yakutia	HOL		19–20th c.	ZIN RAS	Germonpré et al. (2009, 2012), Germonpré, Lázničková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	5
Russia, Kamchatka	HOL		19–20th c.	ZIN RAS	Germonpré et al. (2009, 2012), Germonpré, Lázničková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	4
Russia, Far East	HOL		20th c.	ZIN RAS	Germonpré et al. (2009, 2012), Germonpré, Lázničková-Galetová, et al. (2015), and Germonpré, Sablin, et al. (2015)	5

Abbreviations: HOL, Holocene; ISU, Irkutsk State University, Irkutsk, Russia; LP, Late Pleistocene; MAE RAS, Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Science, Saint-Petersburg, Russia; MZM, Moravian Museum, Anthropos Institute, Brno, the Czech Republic; NHM, Natural History Museum, London, UK; NRM, Swedish Museum of Natural History, Stockholm, Sweden; PM NASU, Palaeontological Museum National Academy of Science of Ukraine, Kiev, Ukraine; RBINS, Royal Belgian Institute of Natural Sciences, Brussels, Belgium; SNM, Natural History Museum of Denmark, Copenhagen, Denmark; ZIN RAS, Zoological Institute of the Russian Academy of Science, Saint-Petersburg, Russia.

carnassial crown length; GWBRC (No. 29), greatest braincase width; GWPAL (No. 34), greatest palatal width; MWPAL (No. 35), minimal palatal width (definition according to von den Driesch, 1976). These measurements have been widely used to differentiate domestic dogs from wild wolves since they describe the morphological changes during the dog domestication process, that is, skull size reduction (TL), broadening of the braincase (GWBRC), shortening and widening of the snout (VL, AL, GWPAL, MWPAL), and dentition reduction (P4CL). TL and GWBRC were taken on the dorsal side of the skull, while the remaining dimensions were measured on the ventral side. All measurements were taken directly from fossils. The measurement error was estimated to be within 5%. There were no missing values in the sample.

2.3 | Raw and size-adjusted (shape) variables

Morphological differences among the large canid groups were evaluated using both raw and size-adjusted measurements. Each variable was ln-transformed prior to all statistical analyses. Size-adjusted (i.e., shape) variables were calculated as residuals from the allometric

regression lines (Jungers et al., 1995) as it was suggested that the morphology of canid skulls varies with their size (Morey, 1992, 2014; Wayne, 1986). Residuals represent size-adjusted morphology as they measure the deviations of actual measurements from the value expected for an average specimen of that particular size (Klingenberg, 1996).

The allometric regression line was fitted by the ordinary least squares regression of each of the original cranial variable (*Y*-axis) on size (*X*-axis) with both variables transformed by natural logarithm (Klingenberg, 1996). The size of each specimen was computed as the geometric mean of all seven raw dimensions (Mosimann & James, 1979). Although reduced major axis regression has been suggested as the preferred allometric technique (cf. Aiello, 1992; Warton, Wright, Falster, & Westoby, 2006), it is not appropriate when residuals from regression line are evaluated as data (Smith, 2009). As our data set consists from multiple groups, analysis of covariance (ANCOVA) tests were used to test for homogeneity of regression slopes (i.e., allometric patterns) among groups. For each variable, common allometric pattern shared by groups was then characterized by the allometric regression line computed from the pooled within-group covariance matrix of the natural ln-transformed data (Thorpe & Leamy, 1983).

2.4 | Univariate and multivariate differences among large canid groups

On a univariate level, differences among the groups on both ln-transformed raw and size-adjusted variables were visualized by jitter plots and tested by analysis of variance (ANOVA) followed by pairwise comparisons between groups using closed Tukey stepwise test procedure, a powerful alternative to conventional Tukey post hoc test (Westfall, 1997). Prior to ANOVA, the assumptions of univariate normality in each group and homogeneity of variance among groups were evaluated using the bootstrapped Shapiro–Wilk's and Levene's tests, respectively.

Multivariate differences in group means were tested by the non-parametric bootstrap method for multivariate analysis of variance (MANOVA) using Wald-type statistic (WTS). In contrast to the ordinary MANOVA, the method is robust to non-normality, heterogeneity of covariance matrices, and small and unbalanced samples (Konietschke, Bathke, Harrar, & Pauly, 2015). Multivariate group differences were then visualized by bgPCA and cluster analysis.

BgPCA is a technique that analyzes multivariate pattern of differences among groups using the projection of the data onto a set of eigenvectors derived from the decomposition of the between-group covariance matrix (bgPCs), which is simply the covariance matrix of the means of the groups (Mitteroecker & Bookstein, 2011). In this context, the discriminant analysis (DA) is usually used because it provides optimal separation between groups by maximizing the between-group to within-group variance ratio (Seetah, Cardini, & Miracle, 2012; Renaud, Dufour, Hardouin, Ledevin, & Auffray, 2015). Distances within a space defined by discriminant axes; however, do not represent original (Euclidean) distances between objects (Mitteroecker & Bookstein, 2011). By contrast, bgPCA does not distort the structure of the data and preserves (Euclidean) distances between objects. It is a simple rotation of the coordination system that does not involve (as in DA) standardization by the pooled within-group variation (Cardini, O'Higgins, & Rohlf, 2019). Furthermore, bgPCA is more robust to assumption violation than DA, it can be computed even when variance–covariance matrices are heterogeneous among groups, variables are intercorrelated, and groups are small in size (Mitteroecker & Bookstein, 2011).

Recent articles (Bookstein, 2019; Cardini et al., 2019) documented that bgPCA may show separation of the groups even when there are no differences among them, especially when an analysis is performed on a large number of variables (p) relative to sample size (i.e., a ratio p/n is large) and when covariance among variables are weak.

In this study, bgPCA was computed from the between-group covariance matrix of all seven cranial variables (both ln-transformed raw and size-adjusted). The very small p/n ratio (0.07) and strong covariance among variables of our data set mitigates problems with spurious group separation observed in bgPCA performed on geometric morphometrics (GMs) data sets characterized by large p/n ratio (cf. Cardini et al., 2019). To visualize the differences among canid groups, two-dimensional convex hull biplots of the scores on the first two bgPCs (bgPC1–2) were drawn.

Cluster analysis provided another perspective on visualizing similarities and differences in a multivariate space. The resulting phenograms were based on Ward's hierarchical clustering of Euclidean distances between specimens computed from all seven ln-transformed raw and size-adjusted variables.

2.5 | Accuracy of large canid groups classification

Following the visualization of group differences using bgPCA and cluster analysis, linear discriminant analysis (DA) was done to examine the accuracy of separation of the large canid groups. Five out of seven variables measured (TL, P4CL, GWBRC, GWPAL, and MWPAL) were included in DA calculation to reduce multicollinearity. DA was carried out on both ln-transformed raw data (DAraw) and size-adjusted data (DAshape). Assumptions of DA were checked to ensure the validity of the model. Multivariate normality in each group was tested by Mardia's test corrected for small samples (Mardia, 1970), the homogeneity of variance–covariance matrices was examined using Box's M test (Dryden & Mardia, 1998), and the existence of multicollinearity among variables was checked by calculating the corrected variance inflation factor (VIF) (Curto & Pinto, 2011).

The overall accuracy of classification was computed as the percentage of correctly classified specimens. All four reference groups were assumed to have equal prior probabilities of classification (i.e., 0.25). The classification accuracy was cross-validated using jackknife (leave-one-out) approach to prevent the overestimation of true accuracy rate. Jackknife is an iterative procedure with the number of iterations being equal to the total sample size ($n = 96$). Within each iteration, one specimen is left out from the sample and its group affiliation is estimated from a DA performed on this smaller sample thus avoiding circular reasoning, that is, classifying an individual based on DA derived from sample which includes the individual itself (for details, see Kovarovic et al., 2011; Galeta et al., 2014).

Specimens were assigned to one of the four a priori reference groups according to the highest posterior probability. In addition to the posterior probability, also the typicality probability was computed. Both probabilities were derived from the (Mahalanobis) distance between specimen and the centroids (i.e., the average specimen) of each reference group. Posterior probability is the probability that the unknown specimen belongs to one of the reference group and is based on the relative distance to the group centroid. Posterior probability assumes that the unknown specimen actually belongs to one of the reference groups, that is, the specimen's posterior probabilities sum to one across the groups (DiGangi & Hefner, 2013). Typicality probability is a measure how many other specimens in a population are expected to be as far or farther from that population centroid than the unknown specimen. It is based on the absolute rather than the relative distance to group centroid (DiGangi & Hefner, 2013). Low typicality probabilities (usually less than 0.01) for all groups indicate that an unknown specimen is either a multivariate outlier or does not belong to any of the reference group included in the analysis (Ousley & Jantz, 2012).

2.6 | Effect of unbalanced samples on accuracy of classification of large canid groups

To explore the effect of the unbalanced composition of our data set on the classification accuracy (overall and by group), three randomization procedures were performed (a) DA with unbalanced-randomized groups, (b) DA with balanced groups, and (c) DA with balanced-randomized groups (cf. Evin et al., 2013). The results of randomization procedures were then compared with the original DA (i.e., DA with unbalanced and non-randomized groups).

Discriminant analysis with unbalanced-randomized groups consisted of generating 1,000 randomized data sets, each with 96 original individuals, whose group affiliations were randomly rearranged among individuals (Evin et al., 2013; Mundry & Sommer, 2007). The purpose of random scrambling was to eliminate group differences. Correct classifications in unbalanced-randomized data sets are only due to chance and they determine inherent (random) variability of classification accuracy for a given sample size and set of variables (Crowley, 1992; Kovarovic et al., 2011). The 95th percentile of classification accuracy in randomized DAs was defined as a random baseline. The random baseline provides an estimate of random chance accuracy; it is the empirical threshold above which classification accuracy is better than chance

(Evin et al., 2013). If the classification accuracy observed in DA based on the original sample is above its random baseline, the original DA performs appreciably better than chance. By contrast, when the observed classification accuracy lies below random baseline then the original DA reflects stochastic variation and does not provide meaningful classification.

Discriminant analysis with balanced groups consisted of generating 1,000 data sets obtained by a random selection of seven individuals from each of the four reference groups (cf. Evin et al., 2013). Every generated data set was perfectly balanced with group sizes equal to the size of the smallest group (Pleistocene wolves). The goal of this randomization experiment was to eliminate the effect of unbalanced groups to the results of DA. If the unbalanced composition of the original data set has no or little effect, classification accuracies observed in the original (unbalanced) DA should be comparable to the classification accuracies in DAs with balanced groups (Evin et al., 2013).

Discriminant analysis with balanced-randomized groups consisted of generating 1,000 balanced data sets with reference group affiliations randomly rearranged among the individuals. A random baseline was again computed as the 95th percentile of classification accuracy in balanced-randomized DAs and gives the empirical threshold above which classification accuracies in balanced data sets are better than chance (Evin et al., 2013). All randomized discriminant analyses were performed using cross-validation, that is, each randomized DA were jackknifed (see above). Statistical analyses were carried out in R 3.5.0 (R Core Team, 2019).

3 | RESULTS

3.1 | Allometric analysis

The tests of homogeneity of allometric slopes among four canid groups provided non-significant results ($p > .05$ for each variable). Although one cannot exclude low power, this suggests that bivariate allometric growth pattern is shared among these groups. The examples of common allometric patterns (TL, P4CL, and MWPAL) are presented in Figure 2. Four out of five cranial measurements used in DA depart from isometry when plotted against the size (i.e., 95% intervals of slopes do not overlap with 1.0). TL and GWBRC are negatively allometric while both GWPAL and MWPAL exhibit positive allometry. Only P4CL is isometric, that is, it maintains geometric similarity among individuals of different sizes. Given the allometric relationship between the majority of cranial measurements and the size of individuals, residuals from
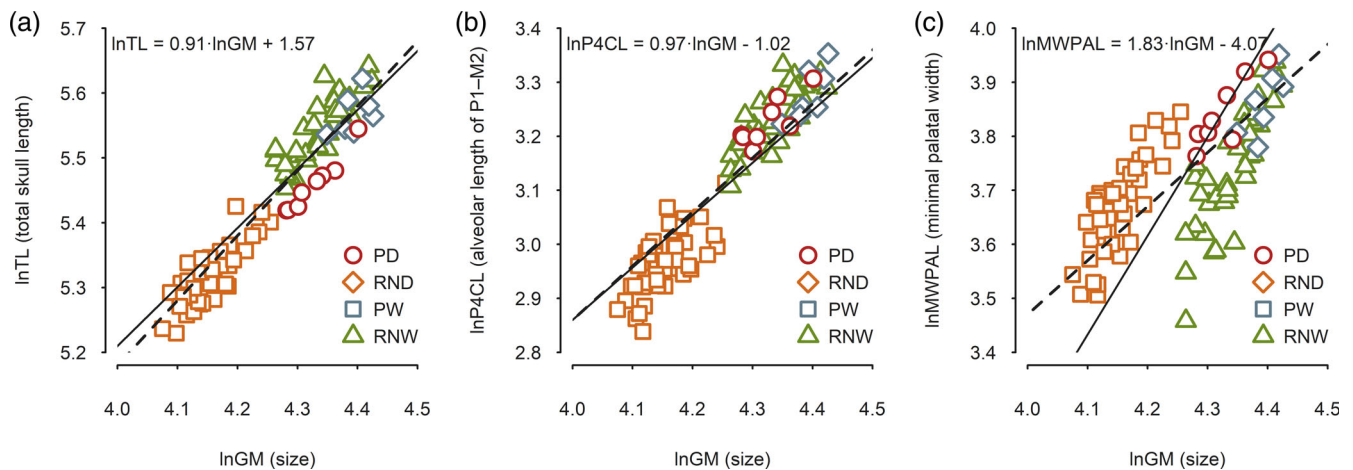
FIGURE 2 Bivariate plots of ln-transformed raw variable against size (geometric mean). (a) Negative allometry, (b) isometry, and (c) positive allometry. Regression line represents the common allometric pattern of four reference groups. Dashed line corresponds to the isometry

common allometric line were taken as size-adjusted variables and used to explore size-dependent shape differences among canids (see Section 2).

3.2 | Univariate differences among large canid groups

The descriptive statistics of raw, ln-transformed raw, size-adjusted, and size variable for all four reference groups is given in Table 2. Distributions of all variables by group are further shown in univariate scatterplots in Figures S1 and S2. All variables satisfy the assumption of homogeneity of variances and normality ($p < .01$) and show significant differences among the means of the four reference groups (ANOVAs, $p < .001$, $df = 3$).

Closed Tukey post hoc test indicates that the overall cranial size (i.e., geometric mean) of Paleolithic dogs is similar to RNWs and lies between RNDs and PWs. Paleolithic dog skull, snout, and tooth row are on average absolutely shorter than those of both recent and Pleistocene wolves (lnTL, lnVL, lnAL). By contrast, craniofacial width dimensions (lnGWBRC, lnGWPAL, and lnMWPAL) and carnassial crown length (lnP4CL) show, with few exceptions, similarities between Paleolithic dogs and wolves (Table 2). There are also differences between PWs and RNWs, with the former group having a greater overall cranial size, a longer viscerocranium and a wider palate than the latter.

Post hoc comparisons of size-adjusted variables (closed Tukey test, Table 2) demonstrate that there are shape differences among Paleolithic dogs and wolves. Paleolithic dogs have, on average, a relatively shorter skull, snout, and tooth row (size-adjusted TL, VL, AL)

and relatively wider palates (size-adjusted MWPAL) than PWs and RNWs. By contrast, carnassial length does not differ between Paleolithic dogs and wolves. Pleistocene wolves lie somewhere between PDs and RNWs and are significantly different from both groups in relative skull and snout length and palate width.

3.3 | Multivariate differences among large canid groups

The mean vectors of the set of seven ln-transformed raw and seven size-adjusted cranial measurements significantly differ among the four canids groups (MANOVA, WTS = 1,850 and 1,200 respectively, $df = 21$, non-parametric bootstrap $p < .001$, 95% confidence interval for effect size $[\eta^2] = 0.49$–0.65 and 0.45–0.61, respectively) and further comparisons show that Paleolithic dogs are different from all three remaining reference groups ($p < .01$).

Figure 3 is a biplot of the bgPCA computed from ln-transformed raw data. Between-group variance explains 76% of the total variance of the data set. The first bgPC (bgPC1) accounts for the majority of the between-group variance visualized in bgPCA (96%) and relates to all seven ln-transformed raw measurements. BgPC2, which explains 3% of the between-group variance, is positively associated with MWPAL and GWPAL and negatively associated mainly with VL and TL. The bgPCA biplot suggests that the sample of Paleolithic dogs fall outside the multivariate distribution of the remaining three reference groups. The position of Paleolithic dogs on the biplot indicates that their VL and TL are intermediate

TABLE 2 Summary statistics for craniometric variables by reference group

	Paleolithic dogs (PDs, n = 8)		Recent northern dogs (RNDs, n = 41)		Pleistocene wolves (PWs, n = 7)		Recent northern wolves (RNWs, n = 40)		Closed Tukey stepwise test procedure					
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	PD vs. RND	PD vs. PW	PD vs. RNW	RND vs. PW	RND vs. RNW	PW vs. RNW
Raw variables														
TL	235.1	10.15	203.8	9.49	262.1	8.31	257.3	12.58	x**	**	**	**	**	
VL	114.1	2.48	99.8	5.62	131.5	3.78	129.3	7.53	x**	**	**	**	**	
AL	82.2	4.61	67.1	3.20	91.2	2.46	86.5	3.15	x**	**	**	**	**	x**
P4CL	25.2	1.15	19.4	1.13	26.6	1.30	25.5	1.43	x**	**		**	**	
GWBRC	67.0	3.45	58.5	2.81	64.7	2.24	65.2	2.63	x**		**	**	**	
GWPAL	82.3	5.10	69.7	4.19	89.0	3.16	79.3	4.97	x**	**		**	**	x**
MWPAL	46.7	3.02	39.4	3.50	47.7	2.86	42.6	4.60	x**		x**	**	**	x**
Ln-transformed raw variables														
lnTL	5.46	0.042	5.32	0.046	5.57	0.031	5.55	0.049	x**	**	**	**	**	
lnVL	4.74	0.022	4.60	0.056	4.88	0.029	4.86	0.059	x**	**	**	**	**	
lnAL	4.41	0.055	4.21	0.047	4.51	0.027	4.46	0.037	x**	**	**	**	**	x**
lnP4CL	3.23	0.045	2.96	0.058	3.28	0.049	3.24	0.057	x**			**	**	
lnGWBRC	4.20	0.051	4.07	0.047	4.17	0.035	4.18	0.041	x**	**		**	**	
lnGWPAL	4.41	0.061	4.24	0.059	4.49	0.036	4.37	0.063	x**	*	**	**	**	x**
lnMWPAL	3.84	0.064	3.67	0.089	3.86	0.060	3.75	0.110	x**		x*	**	**	x*
Size-adjusted (shape) variables														
Adj lnTL	−0.041	0.0093	−0.026	0.0260	0.007	0.0283	0.034	0.0253		**	**	**	**	*
Adj lnVL	−0.064	0.0249	−0.031	0.0363	0.012	0.0201	0.043	0.0353	*	**	**	**	**	*
Adj lnAL	0.011	0.0354	−0.062	0.0276	0.065	0.0168	0.050	0.0285	x**	**	**	**	**	
Adj lnP4CL	0.043	0.0256	−0.050	0.0425	0.029	0.0323	0.038	0.0340	x**			**	**	
Adj lnGWBRC	0.051	0.0428	−0.028	0.0440	−0.005	0.0318	0.019	0.0408	x**	x*	**	**	**	
Adj lnGWPAL	0.003	0.0219	0.054	0.0272	−0.002	0.0233	−0.056	0.0284	**		x**	x**	x**	x**
Adj lnMWPAL	−0.002	0.0378	0.143	0.0533	−0.105	0.0400	−0.128	0.0596	**	x**	x**	x**	x**	x*
Size variable														
GM	75.7	3.17	63.6	2.76	81.0	2.13	77.0	3.46	x**	**	**	**	**	x**
lnGM	4.33	0.041	4.15	0.043	4.39	0.026	4.34	0.045	x**	**	**	**	**	x**

Note: Raw data are reported in millimeters. See Section 2 for abbreviations. x, mean of the first group is larger than mean of the second group in the pairwise comparison.
*Statistically different at 0.05 significance level; **Statistically different at 0.01 significance level.

between recent dogs and wolves but closer to wolves (bgPC1) and that for that particular skull length, Paleolithic dogs tend to have wider palate than wolves (bgPC2). A multivariate pattern provided by bgPCA biplot is reiterated in a cluster analysis phenogram shown in Figure 4. Given the fact that ln-transformed raw data mainly reflect differences in size (cf. high loads on bgPC1 above), phenogram shows two main clusters, RNDs with smaller size versus the remaining three groups with
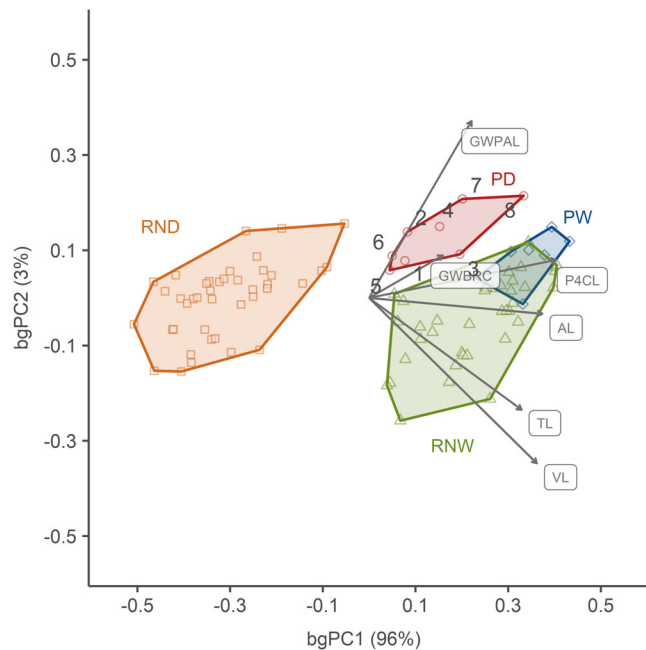
larger-sized specimens. The putative Paleolithic dogs tend to cluster together (with the exception of Eliseevichi 23781, the largest specimen in the PD group) suggesting that despite the similarities in size with RNWs, there appear to be morphological differences between PDs and wolves (cf. bgPC2 above).

Visualization of multivariate differences among the canid groups based on size-adjusted data is provided in bgPCA biplot (Figure 5). Between-group variance again accounts for the majority of the total variance in the data (75%). BgPC1 explains 96% of this between-group variance and contrasts between relative width of palate (MWPAL and GWPAL) and other variables. BgPC2 accounts for 4% of the between-group variance and distinguishes mainly between relative VL and TL and relative GWBRC and P4CL. Paleolithic dog distribution again does not overlap with the other large canid groups suggesting that there are shape differences between Paleolithic dogs, RNDs and both wolf groups in multivariate space. The intermediate position of the Paleolithic dogs along the bgPC1 indicates that their relative palatal width is intermediate between the two wolf groups and RNDs, which have relatively the widest palate among the analyzed groups. The position of the Paleolithic dogs at the positive end of bgPC2 axis reveals that PDs have a relatively short cranium and wide braincase. A similar multivariate pattern is again observed in a cluster analysis phenogram (Figure 6) representing shape differences among groups. There are two main clusters, one consisting of RNDs and the other of the remaining three groups. All putative Paleolithic dogs, however, form a well-defined subcluster and with the exception of a few RNWs and one Pleistocene wolf, they are fairly separated from the wolves. This may be considered as a morphological signal (which is analyzed in detail below) that the putative Paleolithic dogs represent a discrete group with morphological characteristics that distinguish them from both wolf groups.
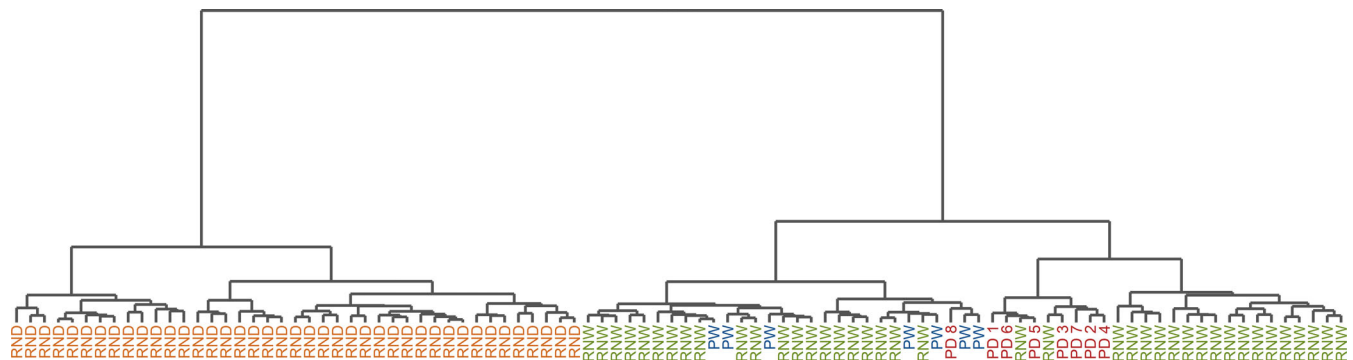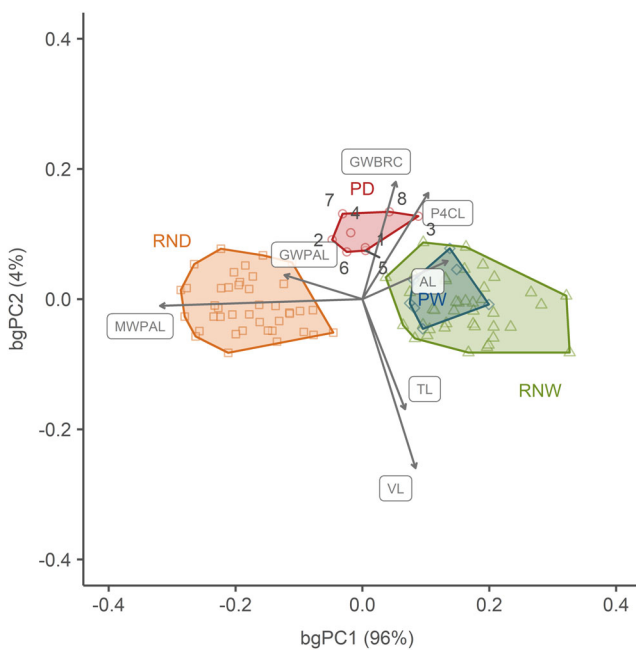

FIGURE 3 Convex hull biplot of four canids groups based on between-groups principal component analysis using seven ln-transformed raw variables. 1, Goyet 2860; 2, Předmostí (−); 3, Předmostí 1060; 4, Předmostí 1063; 5, Mezhirich 4493; 6, Mezin 5490; 7, Eliseevichi 447; 8, Eliseevichi 23781. PDs, Paleolithic dogs; RNDs, recent northern dogs; PWs, Pleistocene wolves; RNWs, recent northern wolves

FIGURE 4 Phenogram based on Ward's hierarchical clustering of Euclidean distances computed from seven ln-transformed raw variables. For abbreviations, see Figure 3

3.4 | Discrimination based on ln-transformed raw data

The tests made suggest that underlying assumptions of discriminant analysis carried out on ln-transformed raw variables (DAraw) are not violated. The entire set of seven cranial measurements was highly intercorrelated, so we limited DAraw to a subset of five variables (TL, P4CL, GWBRC, GWPAL, and MWPAL). Their corrected VIFs ranged between 2.6 and 9.5, that is, below the commonly used threshold of 10 (Montgomery, Peck, & Vining, 2012), which indicates an acceptable level of multicollinearity. Mardia's tests demonstrated that these five variables have multivariate normal distribution

FIGURE 5 Convex hull biplot of four canids groups based on between-groups principal component analysis using seven shape (size-adjusted) variables. For abbreviations, see Figure 3

($p > .4$ for all four groups) and according to Box's M test, variance–covariance matrices are homogenous ($p = .34$).

DAraw provides a very good discrimination (Wilks' lambda 0.032, $p < .001$), which suggests that there are morphological differences among the four canid groups. Jackknife cross-validated classification accuracies for all four reference groups are presented in Table 3. The overall cross-validated classification accuracy is 89.6%. The best classification is achieved in RND group (100%) while the worst prediction (80.0%) is observed in RNWs. The classification accuracy of the putative Paleolithic dogs is high (87.5%) when only one out of eight specimens is incorrectly classified as Pleistocene wolf (Eliseevichi 23781). The remaining seven Paleolithic dogs are unambiguously assigned to PD reference group with the cross-validated posterior probabilities attaining 1.0 (Table 4). High typicality probabilities derived from cross-validated analysis indicate that Paleolithic dogs are typical for their group and are not multivariate outliers.

3.5 | Discrimination based on size-adjusted data

The assumptions of discriminant analysis based on five size-adjusted variables (DAshape) are not violated. The level of multicollinearity is acceptable (corrected VIFs between 2.3 and 9.1), data are multivariate normal (Mardia's tests, $p > .4$ for all four groups) and group variance–covariance matrices are homogenous (Box's M test, $p = .47$).

DAshape discriminates significantly among four canid groups (Wilks' lambda 0.053, $p < .001$). The cross-validated accuracy of correct classification is reported in Table 3. The best cross-validated classification (100%) is carried out in the Paleolithic dog group confirming that the shape of their skull is distinct from the remaining three reference groups. High posterior and typicality
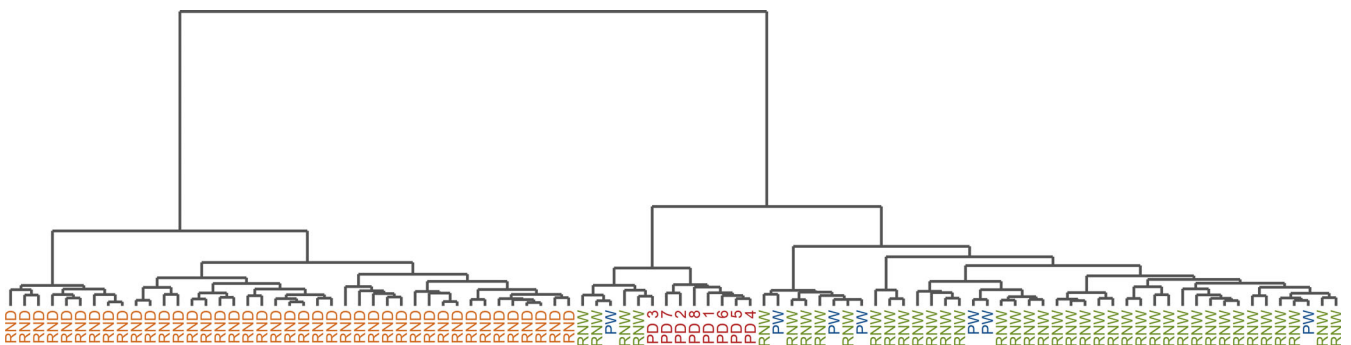

FIGURE 6 Phenogram based on Ward's hierarchical clustering of Euclidean distances computed from seven shape (size-adjusted) variables. For abbreviations, see Figure 3

TABLE 3 Cross-validated classification accuracy (%) and classification table of discriminant analysis based on ln-transformed raw (DAraw) and size-adjusted data (DAshape)

True affiliation	Estimated affiliation											
	DAraw						DAshape					
	Accuracy	PD	RND	PW	RNW	Total	Accuracy	PD	RND	PW	RNW	Total
Paleolithic dogs (PDs)	87.5	7	0	1	0	8	100.0	8	0	0	0	8
Recent northern dogs (RNDs)	100.0	0	41	0	0	41	97.6	1	40	0	0	41
Pleistocene wolves (PWs)	85.7	0	0	6	1	7	71.4	0	0	5	2	7
Recent northern wolves (RNWs)	80.0	1	0	7	32	40	72.5	2	0	9	29	40
Total	89.6	8	41	14	33	96	85.4	11	40	14	31	96

probabilities belonging to the Paleolithic dog group (Table 4) indicate that the classification of the Paleolithic dogs is unambiguous and none of the specimen is multivariate outlier. The lowest classification accuracies based on DAshape show RNWs and PWs (72.5 and 71.4%, respectively). Most of these misclassifications are that specimens from one wolf group are assigned to the other wolf group but there are also two RNWs incorrectly assigned to the Paleolithic dog group.

3.6 | Effect of unbalanced samples on accuracy of classification of large canid groups

Table 5 and Figure 7 compare classification accuracies between the original (unbalanced) sample and the set of 1,000 unbalanced-randomized samples. Both DAraw and DAshape cross-validated classification accuracies computed from the original data set lie above the 95% random baselines. In the DAraw, for example, eight original Paleolithic dogs are classified with the accuracy of 87.5% while the random baseline accuracy is 62.5%. It means that the high classification accuracies of putative Paleolithic dogs in the original DAraw and DAshape (87.5 and 100.0%) cannot be exclusively explained by stochastic variation that is inherent in small and unbalanced samples. The unbalanced composition of the data set might affect the orientation of the discriminant axes but does not seem to strongly bias the results of the classification.

Table 5 and Figure 7 further document that the random variation of the classification accuracy is larger in small- than in large-sized samples. For example, in the DAraw, classification accuracy of eight randomly selected canids assigned as "PDs" attains 62.5% while the accuracy of 41 randomly selected "RNDs" attains only 43.9%.

Classification accuracies in balanced DAs are presented in Table 5 and Figure 8. In balanced DAraw, accuracies ranged between 85.7 and 100.0% for PD and RNW group and between 57.1 and 100.0% for PW and RNW group. Importantly, accuracies in balanced DAs are always comparable with accuracies in DAraw based on the original (unbalanced) data set. It further suggests that unequal sample sizes of our data set do not substantially affect the results of the original DAraw. The same can be applied to the DAshape. Classification rates computed from the original unbalanced DAshape lies well within the limits of accuracies observed in DAshape with balanced groups.

Since DAs with balanced groups are based on very small data sets (28 individuals in total), their results are expected to have large stochastic variation. Random baselines computed from DAs with balanced-randomized groups (57.1% for all groups in both DAraw and DAshape, Table 5 and Figure 8) confirm that accuracies in small data sets may vary considerably from the success rate of 25% predicted by the probability rule for four groups (1/4 = 25%). Nevertheless, classification accuracies in DAs with balanced groups are almost always above these random baselines (except RNW group in DAshape), which indicates that the analysis of balanced data sets performs better than pure chance and provides meaningful results.

4 | DISCUSSION

Univariate and multivariate analyses revealed that there are statistically significant morphological differences between the putative Paleolithic dogs, recent northern dogs, and both wolf groups. Paleolithic dogs, on average, have an absolutely shorter skull and snout than PWs and RNWs (see also Germonpré et al., 2009, 2012, 2017;

TABLE 4 Posterior and typicality probability and estimated group affiliation of Paleolithic dogs (n = 8) based on cross-validated DAraw and DAshape

Specimen	DAraw Posterior probability PD	RND	PW	RNW	Typicality probability PD	RND	PW	RNW	Est. affil.	DAshape Posterior probability PD	RND	PW	RNW	Typicality probability PD	RND	PW	RNW	Est. affil.
Goyet 2860	1.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	PD	1.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	PD
Předmostí (−)	1.00	0.00	0.00	0.00	0.83	0.00	0.02	0.00	PD	0.98	0.02	0.00	0.00	0.37	0.02	0.00	0.00	PD
Předmostí 1060	0.98	0.00	0.01	0.01	0.60	0.00	0.03	0.02	PD	0.97	0.00	0.02	0.01	0.59	0.00	0.04	0.02	PD
Předmostí 1063	1.00	0.00	0.00	0.00	0.97	0.00	0.01	0.01	PD	1.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00	PD
Mezhirich 4493	0.99	0.00	0.00	0.01	0.69	0.00	0.01	0.02	PD	0.97	0.00	0.02	0.01	0.89	0.00	0.10	0.04	PD
Mezin 5490	0.99	0.00	0.00	0.01	0.32	0.00	0.00	0.01	PD	0.93	0.00	0.05	0.02	0.41	0.00	0.05	0.02	PD
Eliseevichi 447	1.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	PD	1.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	PD
Eliseevichi 23781	0.36	0.00	0.62	0.02	0.26	0.00	0.36	0.03	PW	0.93	0.00	0.06	0.01	0.78	0.00	0.16	0.04	PD

Germonpré, Sablin, et al., 2015). Reduction in body and skull size of most animal domesticates has been proposed as a function of changing reproductive strategies under anthropogenic influence (Tchernov & Horwitz, 1991) and most researchers agree that provides good evidence of domestication in zooarcheological assemblages (e.g., Hongo, Pearson, Öksüz, & Ilgezdi, 2009; Marom & Bar-Oz, 2013; Tchernov & Horwitz, 1991). Especially the shortening and widening of the facial part of the skull is judged characteristic of domestication in mammals and has been reported in pigs (Krause-Kyora et al., 2013), cattle (Marom & Bar-Oz, 2013), dogs (Lawrence, 1967; Morey, 1992; Olsen, 1985) and is one of the features of the "domestication syndrome" (Morey & Jeger, 2015; Pendleton et al., 2018; Sánchez-Villagra, Geiger, & Schneider, 2016; Wilkins, Wrangham, & Fitch, 2014). In addition to absolute differences, Paleolithic dogs seem to have, on average, a relatively shorter skull with a relatively shorter and wider snout than the sympatric Pleistocene wolves. Such shape characteristics match well a pattern that has been observed in ancient dog skulls (Benecke, 1994; Lawrence, 1967; Morey, 1992; Nobis, 1986; Olsen, 1985; Sablin & Khlopachev, 2002).

In contrast with the differences in skull size and shape, there are no substantial differences in the mean carnassial tooth length between Paleolithic dogs and wolves. This may be expected, however, because evolutionary reduction in the size of the jugal teeth has been considered to have taken place at a slower rate than the shortening of the snout (Benecke, 1987; Clutton-Brock, 1995; Lawrence, 1967; Morey, 1992).

A particularly important aspect of our analysis is the unbalanced data set composition with small sample size of both fossil reference groups, that is, putative Paleolithic dogs (n = 8) and Pleistocene wolves (n = 7). Given the generally fragmentary nature of the fossil record, however, it is extremely difficult to expand the number of fossils in the analysis. The small sample size is the most limiting factor in most paleobiological studies. Cardini and Elton (2007), for example, showed that about two-thirds of morphometric studies of mammals published between 2001 and 2006 (n = 20) had at least one reference sample equal or smaller than seven (i.e., the size of the smallest reference group [PWs] in our study). Later, Cardini, Seetah, and Barker (2015) considered another list of 40 morphometric articles published between 2008 and 2014 and found than half of these studies included reference samples with less than 10 individuals. As Cardini et al. (2015) and Cardini and Elton (2007) pointed out, small sample size alone should not prevent morphometric analyses from being conducted if the stability of results to unbalanced composition of reference groups is adequately assessed and if limitations are recognized.

TABLE 5 Classification accuracy (%) in DA with unbalanced (original) and balanced data set and their 95% random baseline (%) derived from randomized data sets

| | DAraw | | | | DAshape | | | |
| | Unbalanced groups | | Balanced groups | | Unbalanced groups | | Balanced groups | |
	Accuracy	Random baseline	Accuracy	Random baseline	Accuracy	Random baseline	Accuracy	Random baseline
Paleolithic dogs (n = 8)	87.5	62.5	85.7–100.0	57.1	100.0	62.5	85.7–100.0	57.1
Recent northern dogs (n = 41)	100.0	43.9	85.7–100.0	57.1	97.6	43.9	71.4–100.0	57.1
Pleistocene wolves (n = 7)	85.7	57.1	57.1–100.0	57.1	71.4	57.1	57.1–100.0	57.1
Recent northern wolves (n = 40)	80.0	45.0	57.1–100.0	57.1	72.5	45.0	42.9–100.0	57.1
Total (n = 97)	89.6	35.4	75.0–92.9	39.3	85.4	34.4	75.0–96.4	39.3

Note: Discriminant analysis is based on ln-transformed raw (DAraw) and size-adjusted (DAshape) variables, accuracies are cross-validated. The total sample size in balanced data sets is 28 individuals (seven per group).
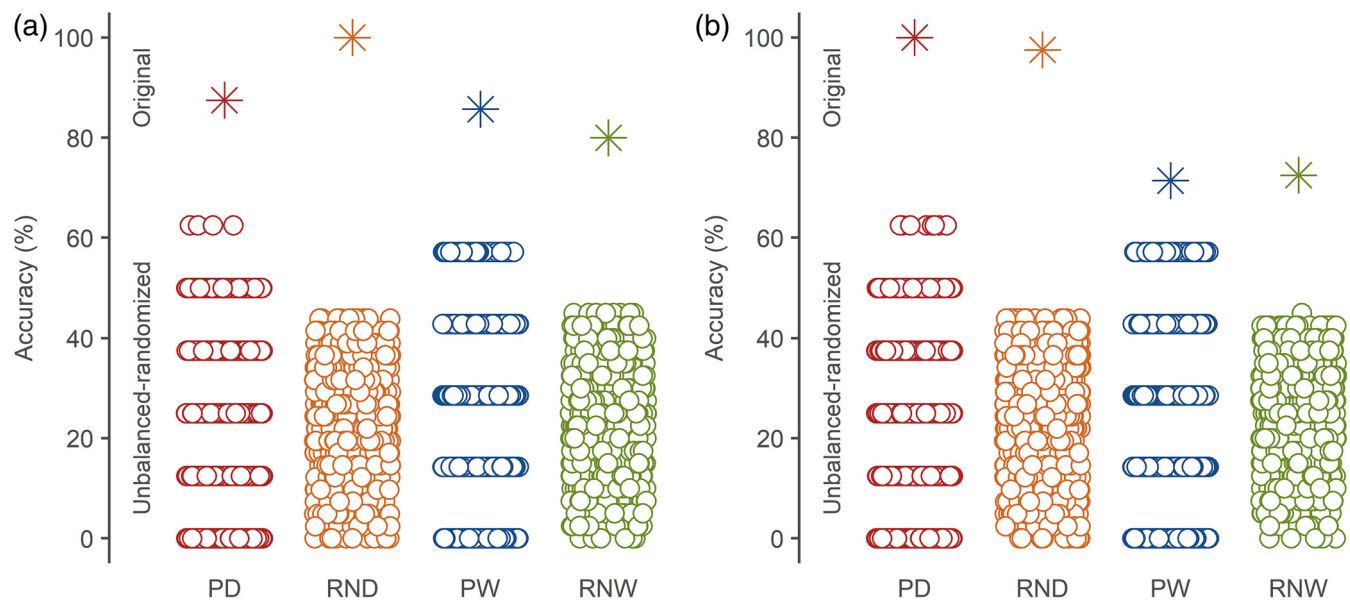
FIGURE 7 Classification accuracy (%) by reference group in the original (stars) and 95% of 1,000 unbalanced-randomized data sets (circles) based on discriminant analysis using five (a) ln-transformed raw and (b) shape (size-adjusted) variables. Random baseline of each group corresponds to the upper limit of accuracies in unbalanced-randomized data sets (circles)

In this study, the effect of small and unbalanced samples on the accuracy of classification is tackled in a systematic way. Firstly, despite the small sample size of two prehistoric groups, our DAs satisfy minimum desirable requirements as the size of the smallest reference group (PWs, $n = 7$) exceed the number of five predictors used in DA (see rules of thumb in Tabachnick & Fidell, 2013). Secondly, assumptions underlying discriminant analysis were examined and found to be not violated. Although tests corrected for small samples were used when

available, small sample sizes may imply low statistical power and tests of DA assumptions may be susceptible to errors. Thirdly, cross-validated accuracy rates were provided as it is known that DAs tend to overfit the data and to overestimate the true accuracy rate (Kovarovic et al., 2011), which may be even aggravated by small and heterogeneous samples.

Fourthly, randomization approach was adopted to assess the stability of the DA results based on small-sized samples. In the biological literature, at least three
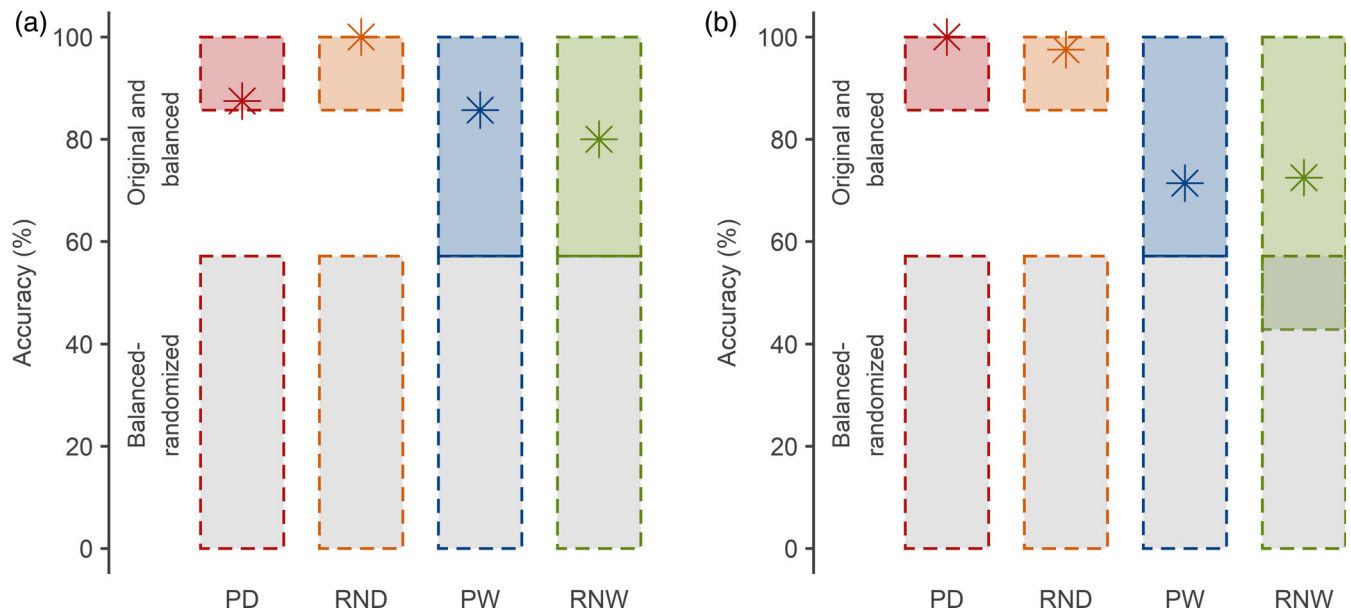
FIGURE 8 Classification accuracy (%) by reference group in the original (stars), 95% of 1,000 balanced data sets (color rectangles) and balanced-randomized data sets (grey rectangles) based on discriminant analysis using five (a) ln-transformed raw and (b) shape (size-adjusted) variables. Random baseline of each group corresponds to the upper limit of accuracies in balanced-randomized data sets

different randomization strategies to measure stability of multivariate analyses to small and unbalanced sample sizes have been discussed: rarefaction, bootstrap, and resampling. In each approach, a large number of random samples are drawn from the original data set. Subsequently, the statistics of interest computed from the (small-sized) original sample is compared to the distribution of this statistic computed from a large number of (small-sized) random samples. Approaches differ in the way the random samples are generated. Rarefaction includes subsampling without replacement from another (large-sized) reference group (Cardini et al., 2015; Cardini & Elton, 2007; Evin et al., 2013; Klingenberg, 2013; Kocovsky, Adams, & Bronte, 2009), bootstrap is sampling with replacement from the original (small-sized) sample (Cardini et al., 2015; Klingenberg, 2013), and resampling involves rearranging the group membership among reference groups of the original data set (both small and large sized) (Evin et al., 2013; Galeta et al., 2014; Kovarovic et al., 2011).

In this study, we adopted a resampling approach to assess the effect of the unbalanced composition of our data set on discriminant analysis results (see Section 2). Discriminant analyses with unbalanced-randomized groups take into account the specific group structure of our data set because they provide results based on the same group sizes as those in the original data set. We documented that the original DAraw and DAshape perform significantly better than 95% of unbalanced-randomized data sets, which suggests that our DAs provide classifications

significantly better than pure chance and that high classification rates are not due to overfit the original data set. Furthermore, we observed little effect of unbalanced composition of our data set on the canid group classification as the accuracies obtained from the DA with perfectly balanced groups are comparable to those obtained in the original DAs. Although analyses based on small and unbalanced samples still need to be interpreted with caution, the high classification accuracy (Table 3) together with high posterior probabilities observed in the original Paleolithic dog group (Table 4) indicate that these individuals do not likely belong to one of the other reference groups. It seems to support the view that the putative Paleolithic dogs may represent a morphologically unique large canid group that is different from both groups of wolves, as well as from recent northern dogs.

The results based on cranial measurements of a small sample of fossil canids correspond well to that obtained from the analysis of much larger number of canid mandibles (Germonpré, Lázničková-Galetová, et al., 2015). Based on the metric and non-metric traits, Germonpré et al. showed that mandibles of putative Paleolithic dogs significantly differs from mandibles of Pleistocene wolves. Their sample consisted from 20 to 36 Paleolithic dogs and 13 to 40 Pleistocene wolves (Germonpré, Lázničková-Galetová, et al., 2015, table 4) and thus provides further support for our conclusions based on the analysis of a smaller number of crania.

Another factor that may account for the morphological differences between putative Paleolithic dogs and

Pleistocene wolves is sexual dimorphism. Boudadi-Maligne and Escarguel (2014, p. 85) explicitly stated that the Paleolithic dog from Goyet most likely represents a female individual and it might be deduced that they consider Paleolithic dogs and Pleistocene wolves to represent females and males of a single fossil species of wolves. Although males of recent grey wolf (*Canis lupus*) are larger than females, wolves exhibit a low level of sexual dimorphism as they tend to live in monogamous pairs (Morris & Brandt, 2014; Van Valkenburgh & Sacco, 2002). For example, skull length of wolf males is on average only 4% larger than that of females (Van Valkenburgh & Sacco, 2002, table 4; Morris & Brandt, 2014, table S2). Other canids (with exception of red fox, *Vulpes vulpes*) showed even lower sexual dimorphism of the skull length (Van Valkenburgh & Sacco, 2002, Table 4). By contrast, Pleistocene wolves (putative males) in our data set have skull length on average 11% (bootstrap confidence interval 7–15%) larger than Paleolithic dogs (putative females). Furthermore, wolf males have relatively broader skulls than females, which is explained as an adaptation to increased bite force in males (Morris & Brandt, 2014). Our results documented, however, that the skull of Pleistocene wolves (putative males) is not broader than the skull of Paleolithic dogs (putative females). Both observations suggest that sexual dimorphism is an unlikely explanation of the morphological differences we observed between Paleolithic dogs and Pleistocene wolves (cf. Germonpré et al., 2012).

Although we detected differences between putative Paleolithic dogs and wolves in multivariate space defined by both bgPCs (Figure 3, Figure 5) and discriminant axes (Table 3), such a pattern is less evident in univariate comparisons of single cranial measurements. For example, only two variables used in DAraw differ significantly between the Paleolithic dogs and the wolves (Table 2). The discrepancies between multivariate and univariate results are, however, not surprising and do not undermine the conclusion that Paleolithic dogs and wolves represent distinct morpho-populations. There may be no single univariate differences between groups but differences may be found when taking into account a whole set of variables (Schumacker, 2015). Multivariate analyses consider the differences on all variables jointly, whereas the univariate analyses treat each variable independently of other variables, that is, they do not include the relation among variables (Stevens, 2012). We suspect that, despite the overlap in single measurements, the unique combination of a relatively shorter skull and a relatively wider palate and braincase distinguishes the putative Paleolithic dogs from wolves.

The sample of eight putative Paleolithic dogs used here is divergent in time and space. They do not form,

however, a completely heterogeneous assemblage as they originated from three specific European regions. Three skulls were found at the same site (Předmostí), and another four in the same region (the Russian Plain, Mezin, Mezhirich, and Eliseevichi sites located at ca. 150–420 km from each other). Only the skull from Goyet site is an isolated case. Naturally, the PD group did not constitute a "population" in the biological sense. With the theoretical exception of the specimens from Předmostí or the specimens from the Russian Plain, they could not interbreed. Temporal and spatial variability of reference samples occurs in most of the retrospective studies. Given the incompleteness of the fossil record, individuals from a broader geographic area and a longer time period are integrated although fossil samples could be highly heterogeneous and estimates of any parameter could be inaccurate due to small sample size or due to mixing individuals from different populations. Our approach is similar to that adopted by many paleoanthropological studies, where all available human remains from entire Europe are routinely combined to form a single EUP and LUP sample (Brewster, Meiklejohn, von Cramon-Taubadel, & Pinhasi, 2014; Holt, 2003; Shaw & Stock, 2013; Sládek et al., 2016) or a single Upper Paleolithic sample (Ruff et al., 2015). BgPCA biplots (Figure 3, Figure 5) also suggest that the putative Paleolithic dogs do not cluster in a multivariate space according to time or geographic origin of specimens. For example, the temporarily and spatially homogenous Paleolithic dogs from Předmostí are dispersed over the entire range of morphological variation of Paleolithic dogs in our sample.

In this study, morphological differences between large canid groups are evaluated within a framework of traditional morphometric (TM) methods, an approach based on linear measurements (Marcus, 1990). TMs have been widely used to distinguish wild from domestic forms in zoological literature (Rowley-Conwy & Zeder, 2014) and many large reference data on size variation between wild and domestic mammals are available (Colledge, Conolly, Dobney, Manning, & Shennan, 2013). In the last 30 years, GMs have been established as a new technique to the study of shape variation of biological objects (for review, see Adams, Rohlf, & Slice, 2004, 2013) and it is considered to be more powerful than TMs (e.g., Adams et al., 2013; Evin et al., 2014). GM studies that compare dog and wolf skulls are so far limited (Drake et al., 2015; Parr et al., 2016; Schmitt & Wallace, 2014) and within very recent studies in this field, TM approach is more common (Boudadi-Maligne & Escarguel, 2014; Crockford & Kuzmin, 2012; Morey, 2014; Napierala & Uerpmann, 2012). Although GM possess indisputable advantages and has been shown to be more sensitive than the TM in discerning morphological differences,

most studies directly comparing GM and TM have revealed that both methods perform equally well in classifying individuals into groups (Breno, Leirs, & Van Dongen, 2011; Franklin, Cardini, Flavel, & Kuliukas, 2012; Macholán, Mikula, & Vohralík, 2008; Mutanen & Pretorius, 2007; Okumura & Araujo, 2014; Santos, Gómez-Olivencia, Arlegi, & Arsuaga, 2017; Viscosi, Lepais, Gerber, & Fortini, 2009). In anticipation of studies based on GM concerning the wolf domestication (Grimm, 2015), our results using TMs yield an accurate and robust classification of large canid groups and may contribute to clarify the first steps in the process of wolf domestication.

Recently, Drake et al. (2015), using 3D GM analysis, reached a contradictory conclusion claiming that EUP dogs are morphologically indistinguishable from wolves. Apart from differences in methodology, the discrepancies between our and Drake et al. results may be attributable to the choice of reference samples. We argue that the putative Paleolithic dogs, dated from the beginning of the domestication process, are more primitive than modern dogs and should be compared with a suitable set of dogs and wolves (Germonpré, Sablin, et al., 2015). Drake et al. (2015), however, do not use Pleistocene Old-World wolves for comparison and many of their modern dogs belong to recent breeds that have been the subject of intensive breeding efforts (cf. Germonpré et al., 2017).

The nature of the Paleolithic dog group cannot be addressed solely by morphometrics. Several lines of evidence must be considered in order to clarify the wolf domestication process. The time frame of the occurrence of the putative Paleolithic dogs at pre-LGM prehistoric sites concurs with the results of a recent study on whole-genome sequence data analyses of prehistoric and recent dogs that place the first steps of the dog domestication process in the Upper Paleolithic, in the period between about 40,000 to 20,000 years ago (Botigué et al., 2017). In addition, other studies of recent and ancient dog patrilines and matrilines suggest several independent domestication processes (Frantz et al., 2016; Smeds, Kojola, & Ellegren, 2019), starting around 29,000 years ago (Smeds et al., 2019). Some of these ancient dog lineages likely went extinct; others were swamped upon by new arrivals, through barter or migration, of new lineages of Paleolithic dogs domesticated in other regions (Germonpré et al., 2017; Germonpré, Láznicková-Galetová, Sablin, & Bocherens, 2018). A few of these lineages survived to this day in modern dogs (Frantz et al., 2016).

Other studies showed (Germonpré et al., 2009, 2012; Germonpré, Bocherens, Láznicková-Galetová, & Sablin, 2020; Sablin & Khlopachev, 2002) that quite a few of the Paleolithic dog skulls were handled peri/postmortem by prehistoric people and that they exhibited

cultural modifications such as perforated braincases (Předmostí, Eliseevichi), cut-marks (Eliseevichi), extraction of teeth (Eliseevichi) and a particular disposal of the specimens (Předmostí, Eliseevichi). In addition, the special social status these canids held in some Upper Paleolithic societies is hinted by the fact that they were found at sites with human remains (Goyet, Předmostí) (Germonpré et al., 2012; Wißing et al., 2019) and/or by their distinct diet and food behavior (Předmostí). The putative Paleolithic dogs were likely not free-roaming but were being fed during life and upon death by Gravettian people (Bocherens et al., 2015; Germonpré et al., 2012, 2017) Furthermore, a recent study by Prassack, DuBois, Láznicková-Galetová, Germonpré, and Ungar (2020) shows that at Předmostí, the two morpho-populations can be separated by diet-related variation in microwear patterning of the second lower molar, with the Paleolithic dog morpho-population showing evidence of greater durophagy than the Pleistocene wolf morpho-population. The shorter, more robust mandibles of the Předmostí incipient dogs (Germonpré, Láznicková-Galetová, et al., 2015) are consistent with the behavior of breaking and consuming more bone (Prassack et al., 2020).

The tight morphological clustering of the putative Paleolithic dogs suggests that a common process is likely related to the similarities within this group. We propose that the distinctiveness of the putative Paleolithic dog group is caused by the incipient domestication that this morpho-population underwent. Combining the evidence of their unique morphology with the presence of their remains at several Upper Paleolithic sites, and the connotation these large canids seem to have had, we propose that the most parsimonious way to view these specimens is as domestic canids, which can but do not need to be the ancestors of the extant dogs but were involved in daily life activities of groups of European Upper Paleolithic foragers (Germonpré et al., 2018, 2020).

5 | CONCLUSIONS

Randomization approach to discriminant analyses employed here strengthen the earlier findings of Germonpré et al. (2009, 2012, 2017) that the putative Paleolithic dogs do not fall within the morphological range of Pleistocene or recent northern Old-World wolves (and recent northern dogs as well) and likely represent a separate Upper Paleolithic canid morpho-population. The results of randomization experiment suggest that the high rates of correct classification of the large canids cannot be exclusively caused by the unbalanced data set composition and small sample sizes of the putative Paleolithic dog

(n = 8) and Pleistocene wolf group (n = 7). Although analyses based on small and unbalanced samples still need to be interpreted with caution, we suggest that the discriminant analysis results reflect the existing morphological differences between the putative Paleolithic dogs and wolf groups rather than the unbalanced composition of the current data set. The putative Paleolithic dogs differs from the Pleistocene and recent northern wolves by the unique combination of a relatively shorter skull and a relatively wider palate and braincase, which can be considered as morphological signs of domestication.

AUTHOR CONTRIBUTIONS

Patrik Galeta: Conceptualization; formal analysis; methodology; software; visualization; writing-original draft; writing-review and editing. **Martina Lázničková-Galetová:** Data curation; funding acquisition; project administration; resources. **Mikhail Sablin:** Data curation; funding acquisition; resources; supervision; validation. **Mietje Germonpré:** Conceptualization; data curation; funding acquisition; investigation; methodology; resources; supervision; writing-original draft; writing-review and editing.

ORCID

Patrik Galeta https://orcid.org/0000-0002-4937-1369
Martina Lázničková-Galetová https://orcid.org/0000-0002-0246-8809
Mietje Germonpré https://orcid.org/0000-0001-8865-0937

REFERENCES

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the "revolution". *The Italian Journal of Zoology*, *71*, 5–16.

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

Aiello, L. C. (1992). Allometry and the analysis of size and shape in human evolution. *Journal of Human Evolution*, *22*, 127–147.

Benecke, N. (1987). Studies on early dog remains from Northern Europe. *Journal of Archaeological Science*, *14*, 31–49.

Benecke, N. (1994). *Archäozoologische Studien zur Entwicklung der Haustierhaltung*. Berlin: Akademie Verlag.

Bocherens, H., Drucker, D. G., Germonpré, M., Lázničková-Galetová, M., Naito, Y. I., Wissing, C., ... Oliva, M. (2015). Reconstruction of the Gravettian food-web at Předmostí I using multi-isotopic tracking (13 C, 15 N, 34 S) of bone collagen. *Quaternary International*, *359*, 211–228.

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

Botigué, L. R., Song, S., Scheu, A., Gopalan, S., Pendleton, A. L., Oetjens, M., ... Veeramah, K. R. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature Communications*, *8*, 16082.

Boudadi-Maligne, M., & Escarguel, G. (2014). A biometric re-evaluation of recent claims for Early Upper Palaeolithic wolf domestication in Eurasia. *Journal of Archaeological Science*, *45*, 80–89.

Breno, M., Leirs, H., & Van Dongen, S. (2011). Traditional and geometric morphometrics for studying skull morphology during growth in *Mastomys natalensis* (Rodentia: Muridae). *Journal of Mammalogy*, *92*, 1395–1406.

Brewster, C., Meiklejohn, C., von Cramon-Taubadel, N., & Pinhasi, R. (2014). Craniometric analysis of European Upper Palaeolithic and Mesolithic samples supports discontinuity at the Last Glacial Maximum. *Nature Communications*, *5*, 4094.

Brown, S. K., Darwent, C. M., & Sacks, B. N. (2013). Ancient DNA evidence for genetic continuity in arctic dogs. *Journal of Archaeological Science*, *40*, 1279–1288.

Cardini, A., & Elton, S. (2007). Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology*, *126*, 121–134.

Cardini, A., & Elton, S. (2011). GeMBiD, a geometric morphometric approach to the study of biological diversity: An example study of the red colobus (Procolobus [Piliocolobus]) species complex. *International Journal of Primatology*, *32*, 377–389.

Cardini, A., O'Higgins, P., & Rohlf, F. J. (2019). Seeing distinct groups where there are none: Spurious patterns from between-group PCA. *Evolutionary Biology*, *46*, 303–316.

Cardini, A., Seetah, K., & Barker, G. (2015). How many specimens do I need? Sampling error in geometric morphometrics: Testing the sensitivity of means and variances in simple randomized selection experiments. *Zoomorphology*, *134*, 149–163.

Clutton-Brock, J. (1995). Origins of the dog: Domestication and early history. In J. Serpell (Ed.), *The domestic dog: Its evolution, behaviour and interactions with people* (pp. 7–20). Cambridge: Cambridge University Press.

Colledge, S., Conolly, J., Dobney, K., Manning, K., & Shennan, S. (2013). *Origins and spread of domestic animals in Southwest Asia and Europe*. Walnut Creek, CA: Left Coast Press.

Crockford, S. J., & Kuzmin, Y. V. (2012). Comments on Germonpré 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 Germonpré, Lázničková-Galetová, and Sablin, Journal of Archaeological Science 39, 2012 "Palaeolithic dog skulls at the Gravettian Předmostí site, The Czech Republic". *Journal of Archaeological Science*, *39*, 2797–2801.

Crowley, P. H. (1992). Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecology and Systematics*, *23*, 405–447.

Curto, J. D., & Pinto, J. C. (2011). The corrected VIF (CVIF). *Journal of Applied Statistics*, *38*, 1499–1507.

DiGangi, E., & Hefner, J. T. (2013). Ancestry estimation. In E. DiGangi & M. K. Moore (Eds.), *Research methods in human skeletal biology* (1st ed., pp. 117–150). Boston, MA: Elsevier.

Drake, A. G., Coquerelle, M., & Colombeau, G. (2015). 3D morphometric analysis of fossil canid skulls contradicts the suggested domestication of dogs during the late Paleolithic. *Scientific Reports*, *5*, 1–8.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Chichester: John Wiley & Sons.

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. *Journal of Archaeological Science*, *40*, 735–743.

Evin, A., Girdland Flink, L., Krause-Kyora, B., Makarewicz, C., Hartz, S., Schreiber, S., ... Dobney, K. (2014). Exploring the complexity of domestication: A response to Rowley-Conwy and Zeder. *World Archaeology*, *46*, 825–834.

Fan, Z., Silva, P., Gronau, I., Wang, S., Armero, A. S., Schweizer, R. M., ... Wayne, R. K. (2016). Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Research*, *26*, 163–173.

Franklin, D., Cardini, A., Flavel, A., & Kuliukas, A. (2012). The application of traditional and geometric morphometric analyses for forensic quantification of sexual dimorphism: Preliminary investigations in a Western Australian population. *International Journal of Legal Medicine*, *126*, 549–558.

Frantz, L. A. F., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., ... Larson, G. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, *252*, 1228–1231.

Galeta, P., Bruzek, J., & Lázničková-Galetová, M. (2014). Is sex estimation from handprints in prehistoric cave art reliable? A view from biological and forensic anthropology. *Journal of Archaeological Science*, *45*, 141–149.

Germonpré, M., Bocherens, H., Lázničková-Galetová, M., & Sablin, M. V. (2020). Could incipient dogs have enhanced differential access to resources among Upper Palaeolithic hunter-gatherers in Europe? In L. Moreau (Ed.), *Social inequality before farming. Multidisciplinary approaches to the study of social organisation in prehistoric and extant hunter-gatherer societies*. Cambridge: McDonald Institute Conversations.

Germonpré, M., Fedorov, S., Danilov, P., Galeta, P., Jimenez, E.-L., Sablin, M., & Losey, R. J. (2017). Palaeolithic and prehistoric dogs and Pleistocene wolves from Yakutia: Identification of isolated skulls. *Journal of Archaeological Science*, *78*, 1–19.

Germonpré, M., Lázničková-Galetová, M., Losey, R. J., Räikkönen, J., & Sablin, M. V. (2015). Large canids at the Gravettian Předmostí site, The Czech Republic: The mandible. *Quaternary International*, *359–360*, 261–279.

Germonpré, M., Laznickova-Galetova, M., & Sablin, M. V. (2012). Palaeolithic dog skulls at the Gravettian Predmosti site, The Czech Republic. *Journal of Archaeological Science*, *39*, 184–202.

Germonpré, M., Lázničková-Galetová, M., Sablin, M. V., & Bocherens, H. (2018). Self-domestication or human control? The upper Palaeolithic domestication of the wolf. In C. Stépanoff & J.-D. Vigne (Eds.), *Hybrid communities: Biosocial approaches to domestication and other trans-species relationships* (pp. 39–64). New York: Routledge.

Germonpré, M., Sablin, M. V., Després, V., Hofreiter, M., Lázničková-Galetová, 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). *Journal of Archaeological Science*, *40*, 786–792.

Germonpré, M., Sablin, M. V., Lázničková-Galetová, M., Després, V., Stevens, R. E., Stiller, M., & Hofreiter, M. (2015). Palaeolithic dogs and Pleistocene wolves revisited: A reply to Morey (2014). *Journal of Archaeological Science*, *54*, 210–216.

Germonpré, M., Sablin, M. V., Stevens, R. E., Hedges, R. E., Hofreiter, M., Stiller, M., & Després, V. R. (2009). Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: Osteometry, ancient DNA and stable isotopes. *Journal of Archaeological Science*, *36*, 473–490.

Grimm, D. (2015). Dawn of the dog. *Science*, *348*, 274–279.

Holt, B. M. (2003). Mobility in Upper Paleolithic and Mesolithic Europe: Evidence from lower limb. *American Journal of Physical Anthropology*, *122*, 200–215.

Hongo, H., Pearson, J., Öksüz, B., & Ilgezdi, G. (2009). The process of ungulate domestication at Cayönü, Southeastern Turkey: A multidisciplinary approach focusing on *Bos* sp. and *Cervus elaphus*. *Anthropozoologica*, *44*, 63–78.

Janssens, L., Perri, A., Crombé, P., Van Dongen, S., & Lawler, D. (2019). An evaluation of classical morphologic and morphometric parameters reported to distinguish wolves and dogs. *J Archaeol Sci Rep*, *23*, 501–533.

Jungers, W. L., Falsetti, A. B., & Wall, C. E. (1995). Shape, relative size, and size-adjustments in morphometrics. *American Journal of Physical Anthropology*, *38*, 137–161.

Klingenberg, C. P. (1996). Multivariate allometry. In L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, & D. E. Slice (Eds.), *Advances in morphometrics* (pp. 23–49). New York: Plenum.

Klingenberg, C. P. (2013). Cranial integration and modularity: Insights into evolution and development from morphometric data. *Hystrix*, *24*, 43–58.

Kocovsky, P. M., Adams, J. V., & Bronte, C. R. (2009). The effect of sample size on the stability of principal components analysis of truss-based fish morphometrics. *Transactions of the American Fisheries Society*, *138*, 487–496.

Konietschke, F., Bathke, A. C., Harrar, S. W., & Pauly, M. (2015). Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis*, *140*, 291–301.

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

Krause-Kyora, B., Makarewicz, C., Evin, A., Flink, L. G., Dobney, K., Larson, G., ... von Wurmb-Schwark, N. (2013). Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nature Communications*, *4*, 2348.

Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y., Peters, J., ... Fredholm, M. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proceedings of the National Academy of Sciences*, *109*, 8878–8883.

Lawrence, B. (1967). Early domestic dogs. *Zeitschrift für Säugetierkunde*, *32*, 44–59.

Macholán, M., Mikula, O., & Vohralík, V. (2008). Geographic phenetic variation of two eastern-Mediterranean non-commensal mouse species, *Mus macedonicus* and *M. cypriacus* (Rodentia: Muridae) based on traditional and geometric approaches to morphometrics. *Zoologischer Anzeiger*, *247*, 67–80.

Marcus, L. F. (1990). Traditional morphometrics. In F. J. Rohlf & F. L. Bookstein (Eds.), *Proceedings of the Michigan morphometrics workshop* (pp. 77–122). Ann Arbor, MI: University of Michigan Museum of Zoology.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.

Marom, N., & Bar-Oz, G. (2013). The prey pathway: A regional history of cattle (*Bos taurus*) and pig (*Sus scrofa*) domestication in the northern Jordan Valley, Israel. *PLoS One*, *8*, e55958.

Mitteroecker, P., & Bookstein, F. (2011). Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology*, *38*, 100–114.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. Hoboken, NJ: John Wiley & Sons.

Morell, V. (2015). How wolf became dog. *Scientific American*, *313*, 48–55.

Morey, D. F. (1992). Size, shape and development in the evolution of the domestic dog. *Journal of Archaeological Science*, *19*, 181–204.

Morey, D. F. (2014). In search of Paleolithic dogs: A quest with mixed results. *Journal of Archaeological Science*, *52*, 300–307.

Morey, D. F., & Jeger, R. (2015). Paleolithic dogs: Why sustained domestication then? *J Archaeol Sci Rep*, *3*, 420–428.

Morris, J. S., & Brandt, E. K. (2014). Specialization for aggression in sexually dimorphic skeletal morphology in grey wolves (*Canis lupus*). *Journal of Anatomy*, *225*, 1–11.

Mosimann, J. E., & James, F. C. (1979). New statistical methods for allometry with applications to Florida red-winged blackbirds. *Evolution*, *33*, 444–459.

Mundry, R., & Sommer, C. (2007). Discriminant function analysis with nonindependent data: Consequences and an alternative. *Animal Behaviour*, *74*, 965–976.

Mutanen, M., & Pretorius, E. (2007). Subjective visual evaluation vs. traditional and geometric morphometrics in species delimitation: A comparison of moth genitalia. *Systematic Entomology*, *32*, 371–386.

Napierala, H., & Uerpmann, H. P. (2012). A "new" Palaeolithic dog from Central Europe. *International Journal of Osteoarchaeology*, *22*, 127–137.

Nobis, G. (1986). Die Wildsäugetiere in der Umwelt des Menschen von Oberkassel bei Bonn und das Domestikationsproblem von Wölfen im Jungpaläolithikum. *Bonner Jahresbücher*, *186*, 367–376.

Okumura, M., & Araujo, A. G. M. (2014). Long-term cultural stability in hunter-gatherers: A case study using traditional and geometric morphometric analysis of lithic stemmed bifacial points from Southern Brazil. *Journal of Archaeological Science*, *45*, 59–71.

Olsen, S. J. (1985). *Origins of the domestic dog: The fossil record*. Tucson, AZ: University of Arizona Press.

Ousley, S. D., & Jantz, R. L. (2012). Fordisc 3 and statistical methods for estimating sex and ancestry. In D. Dirkmaat (Ed.), *A companion to forensic anthropology* (pp. 311–329). Malden, MA: Wiley-Blackwell.

Parr, W. C. H., Wilson, L. A. B., Wroe, S., Colman, N. J., Crowther, M. S., & Letnic, M. (2016). Cranial shape and the modularity of hybridization in dingoes and dogs; hybridization does not spell the end for native morphology. *Evolutionary Biology*, *43*, 171–187.

Pendleton, A. L., Shen, F., Taravella, A. M., Emery, S., Veeramah, K. R., Boyko, A. R., & Kidd, J. M. (2018). Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC Biology*, *16*, 1–21.

Perri, A. (2016). A wolf in dog's clothing: Initial dog domestication and Pleistocene wolf variation. *Journal of Archaeological Science*, *68*, 1–4.

Pidoplichko, I. H. (1998). *Upper Palaeolithic dwellings of mammoth bones in the Ukraine: Kiev-Kirillovskii, Gontsy, Dobranichevka, Mezin and Mezhirich*. Oxford: British Archaeological Reports Limited Series 712.

Pilot, M., Moura, A. E., Okhlopkov, I. M., Mamaev, N. V., Alagaili, A. N., Mohammed, O. B., ... Kopaliani, N. (2019). Global phylogeographic and admixture patterns in grey wolves and genetic legacy of an ancient Siberian lineage. *Scientific Reports*, *9*, 1–13.

Prassack, K. A., DuBois, J., Lázničková-Galetová, M., Germonpré, M., & Ungar, P. S. (2020). Dental microwear as a behavioral proxy for distinguishing between canids at the Upper Paleolithic (Gravettian) site of Předmostí, Czech Republic. *Journal of Archaeological Science*, *115*, 1–10.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Renaud, S., Dufour, A.-B., Hardouin, E. A., Ledevin, R., & Auffray, J.-C. (2015). Once upon multivariate analyses: When they tell several stories about biological evolution. *PLoS One*, *10*, e0132801.

Rowley-Conwy, P., & Zeder, M. (2014). Mesolithic domestic pigs at Rosenhof–or wild boar? A critical re-appraisal of ancient DNA and geometric morphometrics. *World Archaeology*, *46*, 813–824.

Ruff, C. B., Holt, B., Niskanen, M., Sladek, V., Berner, M., Garofalo, E., ... Schuplerova, E. (2015). Gradual decline in mobility with the adoption of food production in Europe. *Proceedings of the National Academy of Sciences*, *112*, 7147–7152.

Sablin, M., & Khlopachev, G. (2002). The earliest ice age dogs: Evidence from Eliseevichi 11. *Current Anthropology*, *43*, 795–799.

Sánchez-Villagra, M. R., Geiger, M., & Schneider, R. A. (2016). The taming of the neural crest: A developmental perspective on the

origins of morphological covariation in domesticated mammals. *Royal Society Open Science*, *3*, 160107.

Santos, E., Gómez-Olivencia, A., Arlegi, M., & Arsuaga, J. L. (2017). Cranial morphological differences within U. deningeri—U. spelaeus lineage: A double traditional and geometric morphometrics approach. *Quaternary International*, *433*(Part A), 347–362.

Schmitt, E., & Wallace, S. (2014). Shape change and variation in the cranial morphology of wild canids (Canis lupus, Canis latrans, Canis rufus) compared to domestic dogs (Canis familiaris) using geometric morphometrics. *International Journal of Osteoarchaeology*, *24*, 42–50.

Schumacker, R. E. (2015). *Using R with multivariate statistics*. Tuscaloosa, AL: SAGE Publications.

Seetah, T. K., Cardini, A., & Miracle, P. T. (2012). Can morphospace shed light on cave bear spatial-temporal variation? Population dynamics of Ursus spelaeus from Romualdova pećina and Vindija (Croatia). *Journal of Archaeological Science*, *39*, 500–510.

Shaw, C. N., & Stock, J. T. (2013). Extreme mobility in the Late Pleistocene? Comparing limb biomechanics among fossil Homo, varsity athletes and Holocene foragers. *Journal of Human Evolution*, *64*, 242–249.

Sládek, V., Ruff, C. B., Berner, M., Holt, B., Niskanen, M., Schuplerová, E., & Hora, M. (2016). The impact of subsistence changes on humeral bilateral asymmetry in terminal Pleistocene and Holocene Europe. *Journal of Human Evolution*, *92*, 37–49.

Smeds, L., Kojola, I., & Ellegren, H. (2019). The evolutionary history of grey wolf Y chromosomes. *Molecular Ecology*, *28*, 2173–2191.

Smith, R. J. (2009). Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, *140*, 476–486.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. New York: Routledge.

Street, M., Napierala, H., & Janssens, L. (2015). The late Paleolithic dog from Bonn-Oberkassel in context. In L. Giemsch & R. W. Schmitz (Eds.), *The Late Glacial burial from Oberkassel revisited. Rheinische Ausgrabungen, Band* (Vol. 72, pp. 253–273). Darmstadt: Philipp von Zabern Verlag.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson Education.

Tchernov, E., & Horwitz, L. K. (1991). Body size diminution under domestication: Unconscious selection in primeval domesticates. *Journal of Anthropological Archaeology*, *10*, 54–75.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., ... Wayne, R. K. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, *342*, 871–874.

Thorpe, R. S., & Leamy, L. (1983). Morphometric studies in inbred and hybrid house mice (Mus sp.): Multivariate analysis of size and shape. *Journal of Zoology*, *199*, 421–432.

van Asch, B., A-b, Z., Oskarsson, M. C. R., Klütsch, C. F. C., Amorim, A., & Savolainen, P. (2013). Pre-Columbian origins of Native American dog breeds, with only limited replacement by European dogs, confirmed by mtDNA analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*, 1–9.

Van Valkenburgh, B., & Sacco, T. (2002). Sexual dimorphism, social behavior, and intrasexual competition in large Pleistocene carnivorans. *Journal of Vertebrate Paleontology*, *22*, 164–169.

Viscosi, V., Lepais, O., Gerber, S., & Fortini, P. (2009). Leaf morphological analyses in four European oak species (Quercus) and their hybrids: A comparison of traditional and geometric morphometric methods. *Plant Biosystems*, *143*, 564–574.

von den Driesch, A. (1976). *A guide to the measurement of animal bones from archaeological sites*. Cambridge, MA: Peabody Museum of Archaeology and Ethnology.

Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews*, *81*, 259–291.

Wayne, R. K. (1986). Cranial morphology of domestic and wild canids: The influence of development on morphological change. *Evolution*, *40*, 243–261.

Westfall, P. H. (1997). Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, *92*, 299–306.

Wilkins, A. S., Wrangham, R. W., & Fitch, W. T. (2014). The "domestication syndrome" in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics*, *197*, 795–808.

Wißing, C., Rougier, H., Baumann, C., Comeyne, A., Crevecoeur, I., Drucker, D. G., ... Bocherens, H. (2019). Stable isotopes reveal patterns of diet and mobility in the last Neandertals and first modern humans in Europe. *Scientific Reports*, *9*, 4433.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.