Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics

Louise Le Meillour, Antoine Zazzo, Joséphine Lesur, Sophie Cersoy, Arul Marie, Matthieu Lebon, David Pleurdeau, Séverine Zirah

► To cite this version:

Louise Le Meillour, Antoine Zazzo, Joséphine Lesur, Sophie Cersoy, Arul Marie, et al.. Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 2018, 511, pp.472-482. 10.1016/j.palaeo.2018.09.013 . hal-02173301

HAL Id: hal-02173301

<https://hal.science/hal-02173301>

Submitted on 20 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics

Louise Le Meillour ^{a, b *}, Antoine Zazzo ^a, Joséphine Lesur ^a, Sophie Cersoy ^c, Arul Marie ^b,
Matthieu Lebon ^d, David Pleurdeau ^d, Séverine Zirah ^b

- a. Unité Archéozoologie, Archéobotanique: Sociétés, pratiques et environnements (AASPE),
Muséum national d'Histoire naturelle, CNRS, CP 56, 55 rue Buffon, 75005 Paris, France
*llemeillour@mnhn.fr zazzo@mnhn.fr jolesur@mnhn.fr
- b. Unité Molécules de Communication et Adaptations des Microorganismes (MCAM), Muséum
national d'Histoire naturelle, CNRS, CP 54, 63 rue Buffon, 75005 Paris, France
armarie@mnhn.fr szirah@mnhn.fr
- c. Centre de Recherche sur la Conservation (CRC), Ministère de la Culture et de la
Communication, Muséum national d'Histoire naturelle, CNRS, CP 21, 36 rue Geoffroy Saint-
Hilaire, 75005 Paris, France
scersoy@mnhn.fr
- d. Unité Histoire Naturelle de l'Homme Préhistorique (HNHP), Muséum national d'Histoire
naturelle, CNRS, 1 rue René Panhard, 75013 Paris, France
lebon@mnhn.fr dpleurd@mnhn.fr

Abstract

The analysis of the skeletal remains of vertebrates in archaeological contexts provides information about human-animal relationship and their environment. Their taxonomic identification based on macroscopic observation is not always possible due to fragmentation and poor preservation. In recent years, proteomics has emerged as an alternative but there is clearly a lack of data in arid environment where diagenesis rapidly affects the integrity of bone proteins. Here, we report the efficiency of three protocols for protein extraction. The protocols used harsh (1 M HCl and 0.6 M HCl) and soft (Tris-EDTA) decalcification agents and were tested on unidentified splinters from the 2000 years-old site of Toteng, Botswana. The preservation of the organic phase was first estimated using attenuated total reflectance Fourier transform infrared spectroscopy and a set of samples with contrasted collagen contents were selected for palaeoproteomics. The extracted proteins were submitted to a bottom-up proteomic approach involving trypsin digestion followed by ultra-high-performance liquid chromatography coupled to mass spectrometry (UHPLC-MS). Our results identify Tris-EDTA buffer as the most suitable decalcification protocol for poorly preserved bones and propose a collagen content threshold of ~ 3% weight content for successful detection of peptides. This approach, combined with biogeographical and chronological repartitions of mammals in Africa allows refining taxonomic attributions for four out of nine splinters, leading to species identification.

Keywords: bone diagenesis, Africa, arid environment, proteomics, ancient collagen, ATR-FTIR

1. Introduction

Bones are among the most abundant material excavated from archaeological sites and thus, potentially the most informative about past societies. Yet, the physical integrity of bone remains is often degraded due to the influence of the environment and burial conditions leading to a large amount of material described as “undetermined” in faunal analyses. New developments in biomolecular methods for the study of ancient remains enabled the archaeologists to use DNA analyses for solving number of archaeological questioning such as phylogenetic relationships (Debruyne et al., 2008; Enk et al., 2011; Hänni et al., 1994) or the genetic impact of species domestication (Bollongino et al., 2006, 2008; Jaenicke-Després et al., 2003; Zeder, 2008). Even though major progresses has been made such as the amplification of the oldest DNA molecules ever identified (Orlando et al., 2013), it is still challenging to extract and amplify this type of biomolecules from arid environment (Kimura et al., 2010; Loosdrecht et al., 2018).

The analysis of protein remains in archaeological samples by palaeoproteomics has become a powerful tool for species identification (Buckley et al., 2008, 2010; Buckley, 2016; van der Sluis et al., 2014), reconstructions of phylogenetic relations between extant and extinct species (Buckley et al., 2011; Cleland et al., 2016, 2015; Welker et al., 2017, 2015a), and sometimes even on unidentified splinters (Welker et al., 2015b). The oldest characterized protein of ostrich eggshell from Tanzania dates back to 3.8 My BP (Demarchi et al., 2016), testifying the potential of palaeoproteomics. But most of the archaeological material analysed by palaeoproteomics comes from temperate climates and not from arid environments. Such environment generates different biochemical and physical constraints on bone matrix, related to differences in temperatures, hydrology, pH or even soil chemical composition.

Often, morphological alterations can come along with biochemical modifications of the organic matrix. Indeed, number of studies focus on both biological and chemical alteration of the bone organic matrix (Collins et al., 2002; Dobberstein et al., 2009; Hedges et al., 1995; Schroeter and Cleland, 2016; Simpson et al., 2016), composed in majority of type I collagen. As the major component of bone extracellular organic matrix, this molecule is formed of two $\alpha 1$ chains and one $\alpha 2$ chain forming a triple helix, further arranged into fibrils (Rich and Crick, 1961). It contains the repetitive sequence Gly – Xaa – Yaa in which Gly is glycine, Xaa can often be either proline (Pro) or hydroxyproline (Hyp) and Yaa any amino acid including Pro or Hyp (Jenkins and Raines, 2002; Kadler et al., 2007; Shoulders and Raines, 2009). The N-and C-terminal regions of the $\alpha 1$ and $\alpha 2$ chains are connected through hydroxylysine cross-links (Eyre et al., 2008). Those structural

characteristics allow the three chains to associate and create empty regions described as gap zones (Xu et al. 2018), where calcium phosphate mineralization occurs, leading to the formation of subsequent hydroxyapatite crystals (Nudelman et al., 2010). Diagenetic processes inherent to archaeological material such as collagen hydrolysis or biochemical modifications (such as deamidation) can increase difficulties associated with biomolecules analyses (Tuross, 2002). Thus, variations occurring in the collagen helix or additional modifications can highly hindered the characterization of biomolecules contained in remains from arid environment.

Despite the increasing number of collected information, no study has ever focused on the characterization of Later Stone Age fauna proteins originating from tropical zones in Africa. To our knowledge, only two papers rely on continental African remains (Demarchi et al., 2016; Prendergast et al., 2017): yet, the first one focused on ostrich eggshells and the second on bone historical material. In Africa, the frequent absence of sufficiently preserved DNA in archaeological material restricted molecular studies (Colson et al., 1997; Gifford-Gonzalez and Hanotte, 2011) and protocols suited for arid environments need to be developed, in order to bring new insights about African faunas recovered in archaeological sites using palaeoproteomics.

The aim of this study was to develop a protein extraction protocol for palaeoproteomics analysis adapted to highly degraded remains from arid environment. The protocols were tested on bone splinters from the 2000 years-old site of Toteng, Botswana. Organic preservation was first estimated using attenuated total reflectance Fourier transform infra-red spectroscopy (ATR-FTIR) and the selected samples underwent three different decalcification protocols. A bottom-up proteomic approach was used to identify digests of the extracted proteins. Finally, protein identification and chrono-biogeographical repartitions of mammals in Africa enabled to propose species attribution of certain of the bone splinters.

2. Material

Fig. 1. Africa map with Botswana localization (dark grey) in Austral Africa and country map with Toteng 1 and Toteng 3 sites localization. **Print in color**

Fig. 2. Pictures of 9 out of the 18 remains selected for this study. The 9 presented are those selected for palaeoproteomics analysis after ATR-FTIR pre-screening. **Print in color**

Analyses were performed on morphologically unidentified splinters from the Later Stone Age sites of Toteng 1 and Toteng 3, Botswana (Fig. 1, left). Discovered in 1991 by A.C. Campbell and collaborators, Toteng 1 and Toteng 3 are open-air sites located West from Ngami Lake basin in Botswana (Fig. 1, right). They are distant only from tens of meters along Ngami Lake tributaries, 934 meters above sea level. First digs revealed a rich sequence (Campbell, 1992; Robbins et al., 1998; Sadr, 1997) with typical Later Stone Age lithic industry and a large spectrum of faunal remains, especially the presence of domestic bovid species, associated with Bambata sherds (Huffman, 1994). In 2003, a new campaign was organized and the two sites delivered new artefacts and ecofacts that enriched the archaeological assemblage (Robbins et al., 2008). Zooarchaeological analyses performed by S. Badenhurst revealed the presence of domestic sheep (*O. aries*) in Later Stone Age levels, directly AMS dated to 2020 ± 40 BP (2σ , Robbins et al., 2005). This early AMS date increased the interest of the site regarding the quest for the earliest domestic caprines in Southern Africa. We accessed the faunal remains ($n=150$) excavated during the 2003 campaign, stored at the National Museum of Botswana, and we temporarily exported them under authorization. Exported remains were examined by one of the authors (JL) and 12 bone fragments and 6 teeth were selected for this study. Most of them were too fragmented and altered by taphonomic processes to allow precise identification (Fig. 2 and Table 1). The identified remains were attributed to bovids and equids.

Sample code	Site	Morphological identification	Anatomical part	Location on the site	Surface alterations visible on the remain
TOT_01	Toteng 1	Unidentified bovid	Ulna diaphysis	Square K, depth 75-80 cm.	Small mammals munching (SMM), rhizomes traces (RT)
TOT_02	Toteng 1	Unidentified bovid	Ulna diaphysis	Square K, depth 75-80 cm.	SMM, RT, burnt?
TOT_03	Toteng 1	Unidentified mammal	Unidentified	Square K, depth 80-85 cm.	SMM, RT
TOT_04	Toteng 1	Medium-sized mammal	Unidentified	Square K, depth 80-85 cm.	SMM, RT, burnt?
TOT_05	Toteng 1	<i>Bos taurus</i> or <i>Syncerus caffer</i>	Upper molar	Square K, depth 90-95 cm.	Large concretions, use wear
TOT_06	Toteng 1	<i>Bos taurus</i> or <i>Syncerus caffer</i>	Upper molar	Square H, depth 115-120 cm.	Large concretions, use wear
TOT_07	Toteng 1	Unidentified	Unidentified	Square L, depth 100-105 cm.	SMM, RT
TOT_09	Toteng 1	Unidentified mammal	Long bone	Square L, depth 100-105 cm.	SMM, RT, crackled
TOT_10	Toteng 1	Small-sized bovid	Lower left molar 3	Square L, depth 110-115 cm.	Concretions, use wear
TOT_11	Toteng 3	<i>Equus</i> sp.	Tooth	Square 7, depth 30_35 cm.	Large concretions, use wear, RT
TOT_12	Toteng 3	Large-sized mammal	Unidentified	Square 7, depth 30_35 cm.	Deep SMM and RT
TOT_13	Toteng 3	Large-sized mammal	Long bone	Square 7, depth 30_35 cm.	Deep SMM, crackled
TOT_14	Toteng 3	Medium-sized bovid	Metapodial	Square 7, depth 30_35 cm.	SMM, RT, crackled
TOT_15	Toteng 3	<i>Equus</i> sp.	Tooth	Square 7, depth 38 cm.	Large concretions, use wear, several pieces
TOT_16	Toteng 3	<i>Equus</i> sp.	Tooth	Square 7, depth 38 cm.	Large concretions, use wear, several pieces
TOT_18	Toteng 3	Large to medium-sized bovid	Proximal phalanx 1	Square 7, depth 30-40 cm.	SMM, RT
TOT_25	Toteng 3	Unidentified mammal	Long bone	Square 7, depth 45-47 cm.	SMM, RT, small diaphysis
TOT_26	Toteng 3	<i>Bos taurus</i> or <i>Syncerus caffer</i>	1 st molar (upper or lower?)	Square 9, depth 45-50 cm.	Large concretions, use wear

Table 1

Archaeological remains selected for the study from the 2003 campaign (Robbins et al., 2005, 2008).

3. Methods

3.1. Sample preparation

The surface of the sample was first cleaned with a diamond drill. Less than 1 mg was sampled for ATR-FTIR analyses and between 20 and 150 mg for palaeoproteomics. Fragments for proteomics were roughly crushed into chunks and then divided into three equal subsamples. The first two subsamples were placed in 2 mL Eppendorf tubes (Protein LoBind, Eppendorf, Hambourg, Germany) while the third one was placed in 5 mL glass tubes.

3.2. ATR-FTIR

Prior to palaeoproteomics analyses, we estimated the organic content of the samples using attenuated total reflectance Fourier transform infra-red (ATR-FTIR) spectroscopy following the method established in Lebon et al., 2016. We applied approximately 0.5 mg of powder on the diamond (Golden Gate Single Reflection Diamond ATR accessory (Specac, France) with KRS-5 lens) of a Vertex 70 FTIR spectrometer (Bruker Optics, France). ATR-FTIR spectra were obtained by accumulation of 64 scans with a resolution of 4 cm⁻¹ in the 4000-400 cm⁻¹ wavenumber range. During acquisition, anvil pressure on sample was adjusted to normalise ν₃PO₄ peaks to an absorbance of 0.5. Resulting data were treated using the OPUS software (Bruker Optics, France) and linear baselines were drawn to measure the area of amide I (between 1710-1590 cm⁻¹) and phosphate ν₃(PO₄) (between 1110-940 cm⁻¹) bands (Fig. S1). Then, the amide I/PO₄ ratio (standard deviation taken into account) was used to estimate nitrogen and collagen weight percent content (%wt) using the equations presented below:

$$\%N \text{ (wt)} = 20.6 \frac{\text{Amide I}}{\text{PO}_4} + 0.31$$

$$\%collagen \text{ (wt)} = 113.13 \frac{\text{Amide I}}{\text{PO}_4} + 1.69$$

More detailed information on data calculation can be obtained in Lebon et al., 2016.

3.3. Demineralization and protein extraction

Fig. 3. Flow chart illustrating the three decalcification protocols employed in this study and further steps of extraction, enzymatic hydrolysis and mass spectrometry analyses. **Print in color**

Three different decalcification protocols were tested (Fig. 3) and an extraction blank performed for each, consisted in the same steps of chemical preparation without any bone powder. The first protocol (Fig. 3, left) follows a previously published method by Buckley et al. (2008). Samples were placed in 1 mL 0.6 M HCl for 4 h at 4 °C. Solutions were replaced every hour after mixing and 10 min of 13 000 rpm centrifugation. After 4 hours, samples were washed 5 times with MilliQ water to reach neutrality.

The second protocol (Fig. 3, center) is an in-house protocol for collagen extraction (Bocherens et al., 2011; Cersoy et al., 2017a). Samples were immersed in 1 mL HCl 1 M for 30 min under 350 rpm agitation at room temperature (ThermoMixer, Eppendorf, Germany). Then, all samples were centrifuged at 13 000 rpm for 10 min and then washed 5 times with MilliQ water to reach neutrality.

In the third protocol (Fig. 3, right), samples were immersed in 1 mL 0.05 M tris(hydroxymethyl)aminomethane (Tris, Sigma Aldrich, Missouri) and 0.5 M ethylenediaminetetraacetic acid (EDTA) buffer, final pH 7.4 as proposed by Balasse and Tresset (2002) and Tuross et al. (1988). The solution was changed daily for 4 days, then every two days.

176 Decalcification was visually and manually checked until a remaining pale-yellow transparent
177 phantom and varied between 1-7 days depending on the sample (Table S1).

178

179 *3.4. Enzymatic hydrolysis*

180

181 Following demineralization, all samples were washed out with MilliQ water. They were then
182 dried under vacuum (SpeedVac Concentrator™, Savant™, ThermoFisher Scientific, Massachusetts).
183 Dried fractions of organic material were solubilized in 100 µL 50 mM ammonium bicarbonate for 3
184 h at 65 °C under 650 rpm agitation (ThermoMixer, Eppendorf, Germany). Supernatants were
185 collected after centrifugation, placed into clean tubes and stored on lab bench until they reached room
186 temperature. Proteomics grade trypsin (Trypsin Gold, Promega, Wisconsin) was added to each
187 sample, 1:100 proportions. They were incubated at 37°C under 650 rpm agitation overnight. Tryptic
188 digestion was stopped adding 1 µL 10% trifluoroacetic acid (TFA, SigmaAldrich, Missouri). Protein
189 concentration of each extract was measured using a microvolume spectrophotometer (NanoVue, GE
190 HealthCare, France) using the pre-set parameters for protein concentration measurement. The
191 concentration values were distributed between 0.5 and 2.3 µg/µL.

192

193 *3.5. MS analysis & data treatment*

194 *3.5.1. Mass spectrometry*

195

196 Tryptic digests were analysed by mass spectrometry (UHPLC-MS). The chromatographic
197 separation was conducted on an Ultimate 3000-RSLC (Thermo Scientific, Massachusetts) using a
198 RSLC Polar Advantage II Acclaim column (2.2 µm, 120 Å, 2.1 × 100 mm, Thermo Scientific,
199 Massachusetts) with a flow rate of 300 µL/min using a mobile phase gradient with A: H₂O + formic
200 acid (FA) 0.1 % and B: acetonitrile (ACN) + FA 0.08 %. Each analysis consisted in 1 µL of sample
201 injected and a blank (5 µL of A-B mix - 20:80) every 5 samples. The mass spectrometer (electrospray
202 ionization, quadrupole, time of flight instrument - ESI Q-TOF, Maxis II ETD, Bruker Daltonics) was
203 used in positive mode on m/z range 200-2200 using an accelerating tension of 3 500 V. Analysis were
204 performed in data dependent auto-MS/MS mode with a cycle time option fixed at 3 sec. and MS
205 acquisition frequency of 2 Hz. Two series of experiments were conducted both with collision induced
206 dissociation (CID): one with low acquisition frequencies to boost MS/MS spectra quality (0.5 Hz for
207 low-abundance ions, 2 Hz for intense ions) and a second with high acquisition frequencies (4 Hz, 16
208 Hz) to maximize the number of MS/MS spectra. Ion selection targeted m/z range 300-2200 with
209 charge states 1⁺ to 5⁺. The mass spectrometer was calibrated with sodium formate clusters at the
210 beginning of each analysis (between 0 and 0.5 min) and data were automatically calibrated.

3.3.2. Data analysis

The LC-MS/MS data were converted to mgf files using DataAnalysis software (Bruker Daltonics, Massachusetts). Database search was carried out against public protein databases (SwissProt and NCBI simultaneously) using the online software Mascot (MatrixScience.com). We restricted the search to « Mammalia (mammals) » protein sequences and allowed 2 missed cleavages for the enzyme. We considered deamidation (N or Q), oxidation (M or P) and phosphorylation (S or T) as variable post-translational modifications (PTM), as these type of PTM were previously reported in the literature for archaeological samples (Buckley and Wadsworth, 2014; Cleland et al., 2015). 1⁺, 2⁺ and 3⁺ peptides were considered with 10 ppm tolerance, MS/MS fragments with 0.1 Da tolerance and the statistic decoy applied to avoid false positive match.

MS/MS data were also analysed using PEAKS DB Studio software (7.5 version, BioInformatics Solutions, Waterloo, Canada, hereafter PEAKS) which permits to propose database search assisted *de novo* sequencing. PEAKS searches (including *de novo*, PEAKS PTM and SPIDER algorithms) were set using the following parameters: 10 ppm precursor mass (monoisotopic) and 0.1 Da fragment ion tolerances; a maximum of 5 PTM per peptide was authorized. PTM set in PEAKS searches were the same as for Mascot, together with the addition of pyroglu from Q and amino acid substitutions. Searches were performed using a restrained collagen database containing sequences described in Table S2. The database was constructed using NCBI protein sequences of alpha 1 and alpha 2 chains of collagen type I (COL1A1 and COL1A2) of the species of interest. The obtained results were filtered using the following parameters: $\leq 1\%$ false discovery rate (FDR) for peptide spectral matches, $\leq 1\%$ FDR and at least one unique peptide for proteins. Finally, peptides that were attributed to a unique species either by Mascot or PEAKS (MS/MS spectra) and all *de novo* peptides were verified using Geneious software (Biomatters, New Zealand) by aligning protein sequences of all species used in the construction of PEAKS restrained database (Table S2). We then made sure that amino acids substitutions between species were not due to any transcription mistake by verifying all codons manually.

4. Results

4.1. ATR-FTIR pre-screening

The collagen contents (collagen wt%) of the samples are summarized in Table 2. They vary between 1.9 and 5.7%, which corresponds to 8.6% to 25.8% of the organic content of modern bone

246 (22.2 %, “*B. taurus* bone (modern reference)” in Table 2). Examples of the best and least preserved
 247 samples spectra are presented in Figure 4. We observed a decrease in the absorbance of the amide I
 248 band absorbance in the archaeological samples compared to the modern sample. Nine samples
 249 covering the entire range of collagen preservation were selected for palaeoproteomics analyses (Table
 250 2).

251
 252 **Fig. 4.** ATR-FTIR spectra of Toteng best preserved (TOT_09, red curve) and least preserved sample (TOT_26, blue
 253 curve) in comparison to modern sample (*B. taurus* bone, yellow curve here). **Print in color**

Sample code	Amide I/PO ₄	%N wt	% Collagen wt
<i>B. taurus</i> bone (modern reference)	0.18	4.04	22.17
TOT_09	0.04	1.04	5.72
TOT_15	0.03	0.98	5.38
TOT_14	0.02	0.65	3.58
TOT_12	0.01	0.61	3.34
TOT_07	0.01	0.57	3.11
TOT_16	0.01	0.57	3.10
TOT_11	0.01	0.54	2.94
TOT_10	0.01	0.50	2.73
TOT_04	0.01	0.49	2.66
TOT_03	0.01	0.47	2.57
TOT_05	0.01	0.45	2.45
TOT_25	0.01	0.45	2.44
TOT_13	0.01	0.44	2.40
TOT_02	0.00	0.41	2.21
TOT_06	0.00	0.40	2.18
TOT_18	0.00	0.37	2.03
TOT_01	0.00	0.36	1.99
TOT_26	0.00	0.35	1.91

254 **Table 2.** ATR-FTIR results. Amide I/ ν 3PO₄ ratio was calculated with OPUS software (Bruker Optics, France) on raw
 255 spectra. N wt% and collagen wt% were estimated after the ratio calculation. Selected samples and their corresponding
 256 collagen wt% for palaeoproteomics analysis are presented in bold black.
 257

258 4. 2. LC-MS analyses
 259

260 Four out of nine remains (TOT_09, TOT_11, TOT_14 and TOT_15) contained identifiable
 261 peptides. The other five (TOT_03, TOT_06, TOT_07, TOT_18 and TOT_26) presented a single
 262 peptide corresponding to trypsin autolysis, VATVSLPR ([M+2H]²⁺, *m/z* 421.75, *rt* 7.6-7.8 min).
 263 Depending on the sample, 4 to 342 MS/MS spectra were generated. All identified peptides and
 264 corresponding proteins are presented for each sample in Table S3. Protocols using Tris-EDTA, 0.6
 265 M HCl and 1 M HCl allowed the identification of 240, 89 and 97 peptides, respectively. Low errors
 266 (between 0 and 3.83 ppm) between the observed and calculated *m/z* for the peptides increased the
 267 confidence in the peptide match. All the identified peptides in the samples were assigned either to
 268 COL1A1 or COL1A2, corresponding to the α 1 and α 2 chains of type I collagen. No other proteins
 269 were identified in any sample. No contamination that could be attributed to peptide signals were
 270 detected in the preparation or the solvent mixture blanks, and no keratins or other human protein were
 271 detected, indicating the reliability of the protocol used in this work.

Fig. 5. Number of assigned peptides per sample regarding the available protein databases by Mascot software. In black, number of assigned peptides using Tris-EDTA; in light grey using 0.6 M HCl and in dark grey using 1 M HCl. Preserved collagen content (collagen wt%) are presented above the corresponding sample.

The analysis suggest that peptide identification depends both on the preservation of the sample and of the decalcification protocol (Fig. 5). Among the samples, TOT_14 yielded the greater number of peptides and TOT_11 the least and all of them presented more identified peptides with Tris-EDTA than with 0.6 and 1 M HCl. PTM were identified in all the protocols with deamidation (NQ) and oxidations (MP) (Table S3). No phosphorylation (ST) was detected. Sequence coverages per sample and protocol range between 3 and 40% (Table 3). Samples decalcified using Tris-EDTA presented a larger coverage (between 18 and 40%) than the ones decalcified using 0.6 and 1 M HCl solutions (respectively ranged between 3-17% and 4-22%). Finally, all samples but TOT_11 presented more than 200 unassigned spectra after Mascot analyses (Table S3), denoting that proteins were hydrolysed at non-tryptic sites or harboured additional PTMs, and/or that the database used was incomplete.

		Tris-EDTA		0.6 M HCl		1 M HCl	
		COL1A1	COL1A2	COL1A1	COL1A2	COL1A1	COL1A2
TOT_09	Mascot	20%	26%	15%	16%	15%	15%
	PEAKS	28%	34%	16%	22%	24%	26%
TOT_11	Mascot	23%	18%	3%	7%	4%	5%
	PEAKS	9%	11%	0%	0%	1%	0%
TOT_14	Mascot	39%	40%	15%	8%	16%	10%
	PEAKS	36%	44%	21%	18%	23%	22%
TOT_15	Mascot	26%	28%	11%	17%	15%	22%
	PEAKS	25%	34%	15%	12%	14%	17%

Table 3. Proteins sequences coverages realised after using Mascot and PEAKS. Each protein sequence coverage is presented by sample (lines) and protocol (columns).

4.3. Species identification – PEAKS results

294 In order to refine Mascot identifications, we analysed raw data using PEAKS software. The
295 use of PEAKS in this study was mainly to take into consideration additional PTMs and amino acid
296 substitution and to compensate the potential lack of information in the databases. PEAKS algorithm
297 is designed to reconstruct database search assisted *de novo* peptides from the MS/MS data and match
298 the generated sequences against a protein database (Zhang et al., 2012). PEAKS PTM and SPIDER
299 tools were used after *de novo* reconstruction. We created a database from the type I collagen
300 sequences reported within bovids, equids and suids in NCBI (Table S2). The limited database was
301 constructed to restrain the search against species that were suspected after morphological
302 identifications (Table 1).

303 Sequence coverages with PEAKS search covered up to 44% of protein sequence and slightly
304 higher 13/24 times (Table 3). Likewise, samples decalcified using Tris-EDTA presented the highest
305 coverage (between 9 and 44%), while coverages were lower when samples were decalcified using
306 HCl (between 0-22% and 0-26% with 0.6 and 1 M HCl, respectively). We observed an increasing
307 number of PTM: in addition to the already identified oxidation (MP) and deamidation (NQ), we
308 observed dimethylation, hydroxylation, phosphorylation (STY) and even P to Y substitution (Fig.
309 S2). Furthermore, the most abundant PTM identified was P hydroxylation (up to 60% of P), as
310 described by Cleland et al. (2015). Species identifications made by Mascot were ascertained and
311 improved with *de novo* peptides as both non-tryptic and modified peptides were reconstructed and
312 assigned to the restrained collagen database.

Fig. 6. MS/MS spectra of COL1A1 distinguishing peptides identified. Letters in bold represent the variant amino acid against other species proteins; letters underlined represent amino acid with PTM. Discriminating ions are represented on each sequence above the corresponding spectrum. Each species sequence is represented by animal shape (horse represents equids, pig represents suids, cow represents bovids except for spectrum (c) where both goat and bovids are represented).

(a) TOT_09, 457... GEPGPTGVQPPGPAGEEGK ...473 ($[M+2H]^{2+}$, m/z 925.9250, rt 6.3 min). Here, 3 PTM have been identified: two P hydroxylations and one Q deamidation. This peptide corresponds to suid COL1A1.

321 (b) TOT_11 and TOT_15, 1068 ... SGDRGEAGPAGPAGPIGPVGAR ... 1089 ([M+3H³⁺] *m/z* 649.3341, *rt* 7.7 min).
 322 No PTM was identified on this peptide. It corresponds to equid COL1A1.
 323 (c) TOT_14, 921 ... AGEVGPPGPPGPAGEK ... 936 ([M+2H²⁺], *m/z* 724.855, *rt* 4.1 min). Here, two P are presenting
 324 hydroxylations. This peptide corresponds to *Ovis aries* COL1A1. **Print in color**

325

326 To ascertain the species identifications, we looked for unique collagen peptides of all family
 327 and species represented in the samples. Even though COL1A2 is commonly used for distinction
 328 because it presents more taxonomic variations, we report here COL1A1 peptides for species and
 329 genera identification. Figure 6 presents the MS/MS spectra corresponding to COL1A1 peptides for
 330 the samples decalcified using Tris-EDTA. Spectrum in Fig. 6a corresponds to a unique peptide of
 331 COL1A1 of suids, identified in sample TOT_09 (long mammal bone). Three PTM are reported here:
 332 two P hydroxylations and one Q deamidation. Similarly, sample TOT_11 (morphologically identified
 333 as an equid tooth) revealed a unique peptide of equid COL1A1 (Fig. 6b). The same peptide was
 334 observed in TOT_15, the second sample identified as an equid both by morphological criteria and
 335 previous search against available databases (this paper, 4.2.). TOT_14, morphologically identified as
 336 a size 2-3 bovid metapodial, was attributed to the species *Ovis aries* (sheep) thanks to 51 unique
 337 peptides of the COL1A1 identified using Mascot (Table S3). The re-examination of this sample's
 338 data highlighted that only one out of the 51 previously identified peptides was indeed specific for the
 339 alpha 1 chain of collagen type I (Fig. 6c) and that 2 amino acids changes can be observed against
 340 other species (Table 4).

341

Species	COL1A1 peptide (921-936)	[M+H] ⁺ <i>m/z</i> monoisotopic	[M+2H] ²⁺ <i>m/z</i>
<i>Ovis aries</i> ; <i>O. aries musimon</i>	AGEVGPPGPPGPAGEK	1416.7118	709.3559
<i>Capra hircus</i>	P.....	1442.7274	722.3637
<i>Bos taurus</i> ; <i>B. mutus</i> ; <i>B. indicus</i> ; <i>B. bison</i>	P.....	1442.7274	722.3637
<i>Equus caballus</i> ; <i>E. asinus</i> ; <i>E. caballus przewalskii</i>	P..A.....	1414.6961	708.3480
<i>Sus scrofa</i> ; <i>S. scrofa domesticus</i>	P..A.....	1414.6961	708.3480
<i>Homo sapiens</i>	P.....	1442.7274	722.3637

342 **Table 4.**
 343 Differences between sheep, goat, bovine, suid and human COL1A1 peptide 921-936.

344

345 **5. Discussion**

346

347 In the context of rare palaeoproteomics analyses of remains originating from Africa, this study
 348 aimed to identify a suitable chemical preparation protocol and then to exploit proteomics data for

349 taxonomic identification of splinters from Botswana. Data analysis allowed to point Tris-EDTA as
350 the best decalcification protocol and to refine taxa attributions of the splinters.

351

352 *5.1. Influence of the decalcification protocols*

353

354 This study explored the use of ATR-FTIR as pre-screening for palaeoproteomics analyses.
355 This method is indeed now used in routine to select samples suitable for radiocarbon dating (Cersoy
356 et al., 2017b; Lebon et al., 2016). Our results show that below 2.9% collagen, no peptide could be
357 identified in the Toteng samples. This threshold is in the same range than that was reported for
358 radiocarbon dating proposed by Cersoy et al. (2017b) and Lebon et al. (2016), set at $\sim 4 \pm 1.2\%$ wt
359 of collagen, suggesting that it could also apply to palaeoproteomics studies. This conclusion,
360 however, is based on a limited number of samples, and more work is needed to further confirmation.
361 Moreover, it is important to remind that the heterogeneity of bone material can introduce biases in
362 the results and that several samplings are to be made to ascertain the reliability of this estimation
363 (Lebon et al., 2016).

364

Sample	Best sequence coverage of COL1A1	Best sequence coverage of COL1A2	Protocol providing the best sequence coverage for COL1	%collagen preserved (ATR-FTIR)
TOT_09	28%	34%	Tris-EDTA	5.72%
TOT_15	25%	34%	Tris-EDTA	5.38%
TOT_14	36%	44%	Tris-EDTA	3.58%
TOT_11	18%	23%	Tris-EDTA	2.94%

365 **Table 5.**

366 Synthetic table of results presenting the best sequence coverages for the two chains of collagen type I, the protocol that
367 allowed best sequence coverage for the whole collagen type I and collagen preservation estimated by ATR-FTIR.

368

369 Table 5 presents a synthesis of the results obtained for the four samples that yielded peptide.
370 Here, the relationship between the sequence coverage and organic phase preservation is not as clear
371 as we might have expected. While the sample with the poorest preservation (TOT_11, 2.94%)
372 presented the lowest sequence coverage (18% and 23% for COL1A1 & A2, respectively) the highest
373 coverage (36% and 44% for COL1A1 & A2, respectively) was obtained for a sample showing
374 intermediate collagen preservation (TOT_14, 3.58%). The triple helix bond by hydroxylysine at their
375 extremity making the collagen structure can be argued as an explanation (Shoulders and Raines,
376 2009). We can assume that the best-preserved samples should present the best collagen integrity. The
377 preservation of collagen crosslinks could limit sequence coverage, as crosslinked peptides would not
378 match to databases. Even if the sequence coverages obtained are not directly comparable to
379 proteomics studies on modern bones (Schroeter et al., 2016), it appears that the protocol used that

380 allowed best sequence coverage is Tris-EDTA. Indeed, recent studies have highlighted the potential
381 of Tris-EDTA for the analysis of bone collagen (Cappellini et al., 2012; Procopio and Buckley, 2017;
382 Sawafuji et al., 2017; Schroeter et al., 2016) and we here point out its efficiency for palaeoproteomics
383 analyses of poorly preserved archaeological bones and teeth.

384 This study also demonstrates the advantage of combining diverse data analysis treatment. The
385 combination of the results obtained with database search (Mascot) and database search assisted *de*
386 *novo* sequencing (PEAKS) allowed increasing peptide reconstruction and protein assignment. The
387 use of database search allowed to estimate which sample presented remaining peptides and the
388 corresponding protein and, in some extent, permitted to propose taxonomic assignments. The use of
389 database search assisted *de novo* sequencing allowed to increase the number of both identified
390 peptides, sequence coverages and to compare the efficiency of the different protocols. A higher
391 number of PTM were observed, especially Q deamidation and P hydroxylation. Glutamine
392 deamidation are often reported in the case of ancient proteins, and are believed to be caused by
393 diagenetic processes (Cleland et al., 2015; Schroeter and Cleland, 2016; van Doorn et al., 2012).
394 Proline hydroxylation is inherent to type I collagen (Kadler et al., 2007) and can also be induced both
395 by acidic treatment and to a lesser extent diagenesis (Cleland et al., 2015; Procopio and Buckley,
396 2017). Identifications with Mascot based on the COL1A2 peptides and the database search assisted
397 *de novo* sequenced peptides of COL1A1 in each positive sample allowed to strongly confirm the
398 identifications previously made and thus giving clear genus and species attributions of the splinters.
399 The results showed that 4 out of 9 splinters presented conserved proteins. The presence of a tryptic
400 peptide and the absence of other proteins in TOT_03, _06, _07, _18 and _26 suggest that they were
401 too degraded and that no organic fraction could be detected by MS analysis. However, and as it was
402 previously demonstrated by Schroeter et al. (2016), there is a possibility for the remaining protein to
403 have been solubilized in decalcification solution (Tris-EDTA, 0.6 M or 1 M HCl). Surprisingly, one
404 sample that presented over ~ 3%wt of collagen preserved (TOT_07, 3.11%collagen wt) did not yield
405 any protein. One explanation would be the overestimate of the collagen content caused by the
406 important content of carbonate (calcite) (Lebon et al., 2016). In any case, the collagen content value
407 of 3.11 is low and, in comparison to the threshold (2.9%), sensibly equivalent taking the error into
408 account (1.4 maximum). Thus, we could argue that a risk of non-identifying preserved proteins does
409 exist in samples with ATR-FTIR values between 2.9 wt% and 3.11 wt% of collagen for
410 palaeoproteomics analyses. In other way, the absence of non-collagenous protein (NCP) in the sample
411 is consistent with what was previously reported in 4 000 years archaeological remains from Cyprus
412 (Buckley and Wadsworth, 2014).

413

5.2. Refining species identification using biogeographical and chronological repartitions of African mammals

Regarding the results presented here, palaeoproteomics analyses allowed the identification of two equids (TOT_11 and TOT_15), one suid (TOT_09) and one sheep (TOT_14). Those attributions remain to genus level, excepted for TOT_14 which was identified to the species level. This can mostly be explained by the lack of protein sequences of African animals in the modern databases. However, the available literature provides guidance regarding the biogeographical and chronological repartition of African mammals (Kingdon, 2015) allowing to refine the proteomics attributions (Table 6). Because Toteng dates from the first century before common era (BCE), the equid remains probably belong to the zebra species *Equus burchellii* (Mammals Species of the World, 2005). Indeed, domestic equids (donkey *E. asinus* and horse *E. caballus*) were brought by Europeans much later, around the XVIIth century (Blench, 2000; Lesur, 2017) and no remains of wild horse (*E. caballus przewalski*) has ever been mentioned on the African continent. Regarding the sample attributed to the suid family (TOT_09) the situation is less clear. Six species coming from four genera of suids do exist nowadays in Africa: *Phacochoerus aethiopicus* and *P. africanus*, *Potamochoerus porcus* and *P. larvatus*, *Hylochoerus meinertzhageni* and *Sus scrofa*. Only the first two genera can correspond to this remain in terms of localization and chronology. Indeed, *Hylochoerus* can be found in western, central and a part of eastern Africa (Grimshaw, 1998; Kingdon, 2015); *Sus scrofa* is endemic of northern Africa (Kingdon, 2015) and its domestic form, *Sus domesticus* was not introduced in Africa before the 5th millenium BCE and arrived in southern Africa mostly by first millennia CE importations (Lesur, 2017; Smith, 2000). In term of species, the geographic distribution thus suggests that the remain could either be attributed to the bushpig *Potamochoerus larvatus* or the common warthog *Phacochoerus africanus* (D'Huart and Grubb, 2001; Grubb, 1993).

Sample code	Morphological attribution	Palaeoproteomics identification	Biogeochronological attribution
TOT_09	Long bone unidentified mammal	Suidae	<i>Phacochoerus</i> or <i>Potamochoerus</i>
TOT_11	Tooth <i>Equus</i> sp.	Equidae	<i>Equus burchellii</i>
TOT_14	Metapodial medium size bovid	<i>Ovis aries</i>	---
TOT_15	Tooth <i>Equus</i> sp.	Equidae	<i>Equus burchellii</i>

Table 6.

Synthetic table of morphological attributions of Toteng remains refined with the combination of palaeoproteomic and biogeographical and chronological data.

6. Conclusion

444

445 This study reports a first attempt at using palaeoproteomics for the species identification of
446 archaeological bones from arid environment of Southern Africa. Estimation of preserved collagen
447 content of bone samples prior to palaeoproteomics analysis using ATR-FTIR permits to propose a
448 threshold of ~ 3%, similar to that previously proposed for radiocarbon dating. The comparison of
449 three decalcification protocols suggests that the use of Tris-EDTA is best suited to poorly preserved
450 samples (less than 5 wt% of collagen preserved as estimated using ATR-FTIR) rather than classical
451 HCl based protocols. Associated with an efficient combination of software for MS data analyses, this
452 study ascertain the use of combined bioinformatics tools for accurate analysis of palaeoproteomics
453 data as previously proposed by Cleland et al. (2015). Both database search (using Mascot search
454 engine) and database search assisted *de novo* sequencing (using PEAKS software) were used to
455 reconstruct and assign remaining peptides. But the use of *de novo* tools and manual examination
456 allowed to ascertain taxonomic identifications. Finally, this study enables to improve morphological
457 identifications for splinters from the Later Stone Age site of Toteng, Botswana. About half of them
458 contained enough proteins to propose a taxonomic attribution. This attribution was further refined by
459 combining proteomics data with biogeographical and chronological distributions of mammals. We
460 then suggest that species identification using palaeoproteomics of remains from Africa should be
461 combined, when possible, with chronological and biogeographical repartitions data. Finally, this
462 study permits the identification of two zebras, one *Phacochoerus* or *Potamochoerus* and one domestic
463 sheep. The latter may potentially have important archaeological implications regarding the
464 introduction of the first domestic caprines in Southern Africa.

465

466 **Acknowledgments**

467

468 This work was supported by a grant from Agence Nationale de la Recherche under the LabEx ANR-
469 10-LABX-0003-BCDiv, in the program « Investissements d'avenir » ANR-11-IDEX-0004-02 and a
470 grant Action Thématique Muséum (ATM) from the Muséum national d'Histoire naturelle under the
471 name « ProtéArch ».

472 The authors would like to thank the National Museum of Botswana for access to the Toteng remains
473 and expressly M. Mvimi for obtaining the exporting permit, without whom this study will not have
474 been possible. The authors also want to thank the Plateau Technique de Spectrométrie de Masse Bio-
475 Organique and the Plateau de Spectrométrie Infrarouge of the MNHN for accessing the instruments
476 and X. Gallet for his help for FTIR analyses.

477 Finally, the authors would like to thank the two anonymous reviewers for their time and comments
478 that contributed to enrich this manuscript.

- 481 Balasse, M., Tresset, A., 2002. Early Weaning of Neolithic Domestic Cattle (Bercy, France)
 482 Revealed by Intra-tooth Variation in Nitrogen Isotope Ratios. *J. Archaeol. Sci.* 29, 853–859.
 483 <https://doi.org/10.1006/jasc.2001.0725>
- 484 Behrensmeyer, A.K., 1978. Taphonomic and ecologic information from bone weathering.
 485 *Paleobiology* 4, 150–162. <https://doi.org/10.1017/S0094837300005820>
- 486 Blench, R.M., 2000. A history of donkeys, wild asses and mules in Africa. *Orig. Dev. Afr. Livest.*
 487 *Archaeol. Genet. Linguist. Ethnogr.* 339–354.
- 488 Bocherens, H., Drucker, D.G., Taubald, H., 2011. Preservation of bone collagen sulphur isotopic
 489 compositions in an early Holocene river-bank archaeological site. *Palaeogeogr. Palaeoclimatol.*
 490 *Palaeoecol., Special Issue: Fossil bones and teeth: preservation or alteration of biogenic*
 491 *compositions?* 310, 32–38. <https://doi.org/10.1016/j.palaeo.2011.05.016>
- 492 Bollongino, R., Edwards, C.J., Alt, K.W., Burger, J., Bradley, D.G., 2006. Early history of
 493 European domestic cattle as revealed by ancient DNA. *Biol. Lett.* 2, 155–159.
- 494 Bollongino, R., Elsner, J., Vigne, J.-D., Burger, J., 2008. Y-SNPs Do Not Indicate Hybridisation
 495 between European Aurochs and Domestic Cattle. *PLoS ONE* 3, e3418.
 496 <https://doi.org/10.1371/journal.pone.0003418>
- 497 Buckley, M., 2016. Species Identification of Bovine, Ovine and Porcine Type 1 Collagen;
 498 Comparing Peptide Mass Fingerprinting and LC-Based Proteomics Methods. *Int. J. Mol. Sci.* 17,
 499 445. <https://doi.org/10.3390/ijms17040445>
- 500 Buckley, M., Collins, M., Thomas-Oates, J., 2008. A method of isolating the collagen (I) $\alpha 2$ chain
 501 carboxyteleopeptide for species identification in bone fragments. *Anal. Biochem.* 374, 325–334.
- 502 Buckley, M., Kansa, S.W., Howard, S., Campbell, S., Thomas-Oates, J., Collins, M., 2010.
 503 Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J.*
 504 *Archaeol. Sci.* 37, 13–20.
- 505 Buckley, M., Larkin, N., Collins, M., 2011. Mammoth and Mastodon collagen sequences; survival
 506 and utility. *Geochim. Cosmochim. Acta* 75, 2007–2016. <https://doi.org/10.1016/j.gca.2011.01.022>
- 507 Buckley, M., Wadsworth, C., 2014. Proteome degradation in ancient bone: Diagenesis and
 508 phylogenetic potential. *Palaeogeogr. Palaeoclimatol. Palaeoecol., Bone and enamel diagenesis:*
 509 *From the crystal to the environment - A tribute to Jean-François Saliège* 416, 69–79.
 510 <https://doi.org/10.1016/j.palaeo.2014.06.026>
- 511 Campbell, A.C., 1992. Southern Okavango integrated development study: archaeological survey of
 512 proposed Maun reservoir.
- 513 Cappellini, E., Jensen, L.J., Szklarczyk, D., Ginolhac, A., da Fonseca, R.A.R., Stafford, T.W.,
 514 Holen, S.R., Collins, M.J., Orlando, L., Willerslev, E., Gilbert, M.T.P., Olsen, J.V., 2012.
 515 Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient
 516 Bone Proteins. *J. Proteome Res.* 11, 917–926. <https://doi.org/10.1021/pr200721u>
- 517 Cersoy, S., Zazzo, A., Lebon, M., Rofes, J., Zirah, S., 2017a. Collagen Extraction and Stable
 518 Isotope Analysis of Small Vertebrate Bones: A Comparative Approach. *Radiocarbon* 59, 679–694.
 519 <https://doi.org/10.1017/RDC.2016.82>
- 520 Cersoy, S., Zazzo, A., Rofes, J., Tresset, A., Zirah, S., Gauthier, C., Kaltnecker, E., Thil, F.,
 521 Tisnerat-Laborde, N., 2017b. Radiocarbon dating minute amounts of bone (3–60 mg) with
 522 ECHoMICADAS. *Sci. Rep.* 7, 7141. <https://doi.org/10.1038/s41598-017-07645-3>
- 523 Cleland, T.P., Schroeter, E.R., Feranec, R.S., Vashishth, D., 2016. Peptide sequences from the first
 524 *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proc R*
 525 *Soc B* 283, 20160593. <https://doi.org/10.1098/rspb.2016.0593>
- 526 Cleland, T.P., Schroeter, E.R., Schweitzer, M.H., 2015. Biologically and diagenetically derived
 527 peptide modifications in moa collagens. *Proc R Soc B* 282, 20150015.
 528 <https://doi.org/10.1098/rspb.2015.0015>
- 529 Collins, M.J., Nielsen-Marsh, C.M., Hiller, J., Smith, C.I., Roberts, J.P., Prigodich, R.V., Wess,
 530 T.J., Csapò, J., Millard, A.R., Turner-Walker, G., 2002. The survival of organic matter in bone: a
 531 review. *Archaeometry* 44, 383–394. <https://doi.org/10.1111/1475-4754.t01-1-00071>

Colson, I., Baird, J., Vercauteren, M., Sykes, B.C., Hedges, R.E.M., 1997. The preservation of ancient DNA and bone diagenesis. *Anc. Biomol.* 1, 109–117.

Debruyne, R., Chu, G., King, C.E., Bos, K., Kuch, M., Schwarz, C., Szpak, P., Gröcke, D.R., Matheus, P., Zazula, G., Guthrie, D., Froese, D., Buigues, B., de Marliave, C., Flemming, C., Poinar, D., Fisher, D., Southon, J., Tikhonov, A.N., MacPhee, R.D.E., Poinar, H.N., 2008. Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths. *Curr. Biol.* 18, 1320–1326. <https://doi.org/10.1016/j.cub.2008.07.061>

Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B.M., Jersie-Christensen, R.R., Olsen, J.V., Haile, J., Thomas, J., Marean, C.W., Parkington, J., Presslee, S., Lee-Thorp, J., Ditchfield, P., Hamilton, J.F., Ward, M.W., Wang, C.M., Shaw, M.D., Harrison, T., Domínguez-Rodrigo, M., MacPhee, R.D., Kwekason, A., Ecker, M., Horwitz, L.K., Chazan, M., Kröger, R., Thomas-Oates, J., Harding, J.H., Cappellini, E., Penkman, K., Collins, M.J., 2016. Protein sequences bound to mineral surfaces persist into deep time. *eLife* 5, e17092. <https://doi.org/10.7554/eLife.17092>

D'Huart, J.-P., Grubb, P., 2001. Distribution of the common warthog (*Phacochoerus africanus*) and the desert warthog (*Phacochoerus aethiopicus*) in the Horn of Africa. *Afr. J. Ecol.* 39, 156–169. <https://doi.org/10.1046/j.0141-6707.2000.00298.x>

Dobberstein, R.C., Collins, M.J., Craig, O.E., Taylor, G., Penkman, K.E.H., Ritz-Timme, S., 2009. Archaeological collagen: Why worry about collagen diagenesis? *Archaeol. Anthropol. Sci.* 1, 31–42. <https://doi.org/10.1007/s12520-009-0002-7>

Enk, J., Devault, A., Debruyne, R., King, C.E., Treangen, T., O'Rourke, D., Salzberg, S.L., Fisher, D., MacPhee, R., Poinar, H., 2011. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biol.* 12, R51. <https://doi.org/10.1186/gb-2011-12-5-r51>

Eyre, D.R., Weis, M.A., Wu, J.-J., 2008. Advances in collagen cross-link analysis. *Methods, Methods in Extracellular Matrix Research* 45, 65–74. <https://doi.org/10.1016/j.ymeth.2008.01.002>

Gifford-Gonzalez, D., Hanotte, O., 2011. Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. *J. World Prehistory* 24, 1–23. <https://doi.org/10.1007/s10963-010-9042-2>

Grimshaw, J.M., 1998. The giant forest hog *Hylochoerus meinertzhageni* in Tanzania records rejected. *Mammalia* 62, 123–125.

Grubb, P., 1993. The afrotropical suids (*Phacochoerus*, *Hylochoerus* and *Potamochoerus*). Taxonomy and description. Pigs Peccaries Hippos Status Surv. Conserv. Plan IUCN Gland Switz.

Hänni, C., Laudet, V., Stehelin, D., Taberlet, P., 1994. Tracking the origins of the cave bear (*Ursus spelaeus*) by mitochondrial DNA sequencing. *Proc. Natl. Acad. Sci.* 91, 12336–12340.

Hedges, R.E.M., Millard, A.R., Pike, A.W.G., 1995. Measurements and Relationships of Diagenetic Alteration of Bone from Three Archaeological Sites. *J. Archaeol. Sci.* 22, 201–209. <https://doi.org/10.1006/jasc.1995.0022>

Huffman, T.N., 1994. Toteng pottery and the origins of Bambata. *South. Afr. Field Archaeol.* 3, 3–9.

Jaenicke-Després, V., Buckler, E.S., Smith, B.D., Gilbert, M.T.P., Cooper, A., Doebley, J., Pääbo, S., 2003. Early Allelic Selection in Maize as Revealed by Ancient DNA. *Science* 302, 1206–1208. <https://doi.org/10.1126/science.1089056>

Jenkins, C.L., Raines, R.T., 2002. Insights on the conformational stability of collagen. *Nat. Prod. Rep.* 19, 49–59. <https://doi.org/10.1039/A903001H>

Kadler, K.E., Baldock, C., Bella, J., Boot-Handford, R.P., 2007. Collagens at a glance. *J. Cell Sci.* 120, 1955–1958. <https://doi.org/10.1242/jcs.03453>

Kimura, B., Marshall, F.B., Chen, S., Rosenbom, S., Moehlman, P.D., Tuross, N., Sabin, R.C., Peters, J., Barich, B., Yohannes, H., Kebede, F., Teclai, R., Beja-Pereira, A., Mulligan, C.J., 2010. Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proc. R. Soc. Lond. B Biol. Sci.* [rsbp20100708](https://doi.org/10.1098/rspb.2010.0708). <https://doi.org/10.1098/rspb.2010.0708>

Kingdon, J., 2015. The Kingdon field guide to African mammals. Bloomsbury Publishing.

Lebon, M., Reiche, I., Gallet, X., Bellot-Gurlet, L., Zazzo, A., 2016. Rapid Quantification of Bone Collagen Content by ATR-FTIR Spectroscopy. *Radiocarbon* 1–15.

Lesur, J., 2017. Et la gazelle devint chèvre: pré-histoires africaines d’hommes et d’animaux, Presses Universitaires du Midi. ed, Sites et Cités d’Afrique. Toulouse.

Loosdrecht, M. van de, Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., Hajraoui, M.A.E., Amzazi, S., Hublin, J.-J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W., Jeong, C., Krause, J., 2018. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* eaar8380. <https://doi.org/10.1126/science.aar8380>

Mammals Species of the World. A Taxonomic and Geographic Reference, John Hopkins University Press. ed, 2005. . Don E. Wilson & DeAnn M. Reeder.

Nudelman, F., Pieterse, K., George, A., Bomans, P.H.H., Friedrich, H., Brylka, L.J., Hilbers, P.A.J., de With, G., Sommerdijk, N.A.J.M., 2010. The role of collagen in bone apatite formation in the presence of hydroxyapatite nucleation inhibitors. *Nat. Mater.* 9, 1004–1009. <https://doi.org/10.1038/nmat2875>

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P.L.F., Fumagalli, M., Vilstrup, J.T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C.D., Vinther, J., Dolocan, A., Stenderup, J., Velazquez, A.M.V., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G.D., Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J.F., Weinstock, J., Gregersen, K., Røed, K.H., Eisenmann, V., Rubin, C.J., Miller, D.C., Antczak, D.F., Bertelsen, M.F., Brunak, S., Al-Rasheid, K.A.S., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M.T.P., Kjær, K., Sicheritz-Ponten, T., Jensen, L.J., Olsen, J.V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., Willerslev, E., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74–78. <https://doi.org/10.1038/nature12323>

Prendergast, M.E., Buckley, M., Crowther, A., Frantz, L., Eager, H., Lebrasseur, O., Hutterer, R., Hulme-Beaman, A., Neer, W.V., Douka, K., Veall, M.-A., Morales, E.M.Q., Schuenemann, V.J., Reiter, E., Allen, R., Dimopoulos, E.A., Helm, R.M., Shipton, C., Mwebi, O., Denys, C., Horton, M., Wynne-Jones, S., Fleisher, J., Radimilahy, C., Wright, H., Searle, J.B., Krause, J., Larson, G., Boivin, N.L., 2017. Reconstructing Asian faunal introductions to eastern Africa from multi-proxy biomolecular and archaeological datasets. *PLOS ONE* 12, e0182565. <https://doi.org/10.1371/journal.pone.0182565>

Procopio, N., Buckley, M., 2017. Minimizing Laboratory-Induced Decay in Bone Proteomics. *J. Proteome Res.* 16, 447–458. <https://doi.org/10.1021/acs.jproteome.6b00564>

Rich, A., Crick, F.H.C., 1961. The molecular structure of collagen. *J. Mol. Biol.* 3, 483–IN4. [https://doi.org/10.1016/S0022-2836\(61\)80016-8](https://doi.org/10.1016/S0022-2836(61)80016-8)

Robbins, L.H., Campbell, A.C., Murphy, M.L., Brook, G.A., Liang, F., Skaggs, S.A., Srivastava, P., Mabuse, A.A., Badenhorst, S., 2008. Recent archaeological research at Toteng, Botswana : Early domesticated livestock in the Kalahari. *J. Afr. Archaeol.* 131–149.

Robbins, L.H., Campbell, A.C., Murphy, M.L., Brook, G.A., Srivastava, P., Badenhorst, S., 2005. The Advent of Herding in Southern Africa: Early AMS Dates on Domestic Livestock from the Kalahari Desert. *Curr. Anthropol.* 46, 671–677. <https://doi.org/10.1086/432748>

Robbins, L.H., Murphy, M.L., Campbell, A.C., Brook, G.A., Reid, D.M., Haberyan, K.A., Downey, W.S., 1998. Test excavations and reconnaissance palaeoenvironmental work at Toteng, Botswana. *South Afr. Archaeol. Bull.* 53, 125–132.

Sadr, K., 1997. Kalahari archaeology and the Bushman debate. *Curr. Anthropol.* 38, 104–112.

Sawafuji, R., Cappellini, E., Nagaoka, T., Fotakis, A.K., Jersie-Christensen, R.R., Olsen, J.V., Hirata, K., Ueda, S., 2017. Proteomic profiling of archaeological human bone. *R. Soc. Open Sci.* 4. <https://doi.org/10.1098/rsos.161004>

Schroeter, E.R., Cleland, T.P., 2016. Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Commun. Mass Spectrom.* 30, 251–255.

<https://doi.org/10.1002/rcm.7445>
 Schroeter, E.R., DeHart, C.J., Schweitzer, M.H., Thomas, P.M., Kelleher, N.L., 2016. Bone protein
 “extractomics”: comparing the efficiency of bone protein extractions of *Gallus gallus* in tandem
 mass spectrometry, with an eye towards paleoproteomics. *PeerJ* 4, e2603.
<https://doi.org/10.7717/peerj.2603>
 Shoulders, M.D., Raines, R.T., 2009. Collagen Structure and Stability. *Annu. Rev. Biochem.* 78,
 929–958. <https://doi.org/10.1146/annurev.biochem.77.032207.120833>
 Simpson, J.P., Penkman, K.E.H., Demarchi, B., Koon, H., Collins, M.J., Thomas-Oates, J., Shapiro,
 B., Stark, M., Wilson, J., 2016. The effects of demineralisation and sampling point variability on
 the measurement of glutamine deamidation in type I collagen extracted from bone. *J. Archaeol. Sci.*
 69, 29–38. <https://doi.org/10.1016/j.jas.2016.02.002>
 Smith, A.B., 2000. The origins of domesticated animals of Southern Africa, in: Blench, R.M.,
 MacDonald, K.C. (Eds.), *The Origins and Development of African Livestock: Archaeology,*
Genetics, Linguistics and Ethnography. London, pp. 222–238.
 Tuross, N., 2002. Alterations in fossil collagen. *Archaeometry* 44, 427–434.
 Tuross, N., Fogel, M.L., Hare, P.E., 1988. Variability in the preservation of the isotopic
 composition of collagen from fossil bone. *Geochim. Cosmochim. Acta* 52, 929–935.
 van der Sluis, L.G., Hollund, H.I., Buckley, M., De Louw, P.G.B., Rijdsdijk, K.F., Kars, H., 2014.
 Combining histology, stable isotope analysis and ZooMS collagen fingerprinting to investigate the
 taphonomic history and dietary behaviour of extinct giant tortoises from the Mare aux Songes
 deposit on Mauritius. *Palaeogeogr. Palaeoclimatol. Palaeoecol., Bone and enamel diagenesis: From*
the crystal to the environment - A tribute to Jean-François Saliège 416, 80–91.
<https://doi.org/10.1016/j.palaeo.2014.06.003>
 van Doorn, N.L., Wilson, J., Hollund, H., Soressi, M., Collins, M.J., 2012. Site-specific
 deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass*
Spectrom. 26, 2319–2327. <https://doi.org/10.1002/rcm.6351>
 Welker, F., Collins, M.J., Thomas, J.A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S.T.,
 Reguero, M., Gelfo, J.N., Kramarz, A., Burger, J., Thomas-Oates, J., Ashford, D.A., Ashton, P.D.,
 Rowsell, K., Porter, D.M., Kessler, B., Fischer, R., Baessmann, C., Kaspar, S., Olsen, J.V., Kiley,
 P., Elliott, J.A., Kelstrup, C.D., Mullin, V., Hofreiter, M., Willerslev, E., Hublin, J.-J., Orlando, L.,
 Barnes, I., MacPhee, R.D.E., 2015a. Ancient proteins resolve the evolutionary history of Darwin’s
 South American ungulates. *Nature* 522, 81–84. <https://doi.org/10.1038/nature14249>
 Welker, F., Smith, G.M., Hutson, J.M., Kindler, L., Garcia-Moreno, A., Villaluenga, A., Turner, E.,
 Gaudzinski-Windheuser, S., 2017. Middle Pleistocene protein sequences from the rhinoceros genus
Stephanorhinus and the phylogeny of extant and extinct Middle/Late Pleistocene *Rhinocerotidae*.
PeerJ 5, e3033. <https://doi.org/10.7717/peerj.3033>
 Welker, F., Soressi, M., Rendu, W., Hublin, J.-J., Collins, M., 2015b. Using ZooMS to identify
 fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *J.*
Archaeol. Sci. 54, 279–286. <https://doi.org/10.1016/j.jas.2014.12.010>
 Xu, Z., Zhao, W., Wang, Z., Yang, Y., Sahai, N., 2018. Structure analysis of collagen fibril at
 atomic-level resolution and its implications for intra-fibrillar transport in bone biomineralization.
Phys. Chem. Chem. Phys. 20, 1513–1523. <https://doi.org/10.1039/C7CP05261H>
 Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins,
 diffusion, and impact. *Proc. Natl. Acad. Sci.* 105, 11597–11604.
<https://doi.org/10.1073/pnas.0801317105>
 Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma,
 B., 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate
 peptide identification. *Mol. Cell. Proteomics MCP* 11, M111.010587.
<https://doi.org/10.1074/mcp.M111.010587>