

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Fabrice Bray^{a,*}, Stéphanie Flament^a, Grégory Abrams^{b,c}, Dominique Bonjean^{b,d}, Kévin Di Modica^b, Christian Rolando^{a,e}, Caroline Tokarski^f, Patrick Auguste^g

^a Univ. Lille, CNRS, USR 3290 - MSAP - Miniaturisation pour la Synthèse, l'Analyse et la Protéomique, F-59000 Lille, France

^b Scladina Cave Archaeological Centre, Rue Fond des Vaux 339d, 5300, Sclayn, Belgium

^c Faculty of Archaeology, Leiden University, Van Steenis Building, Einsteinweg 2, 2333 CC Leiden, The Netherlands

^d Department of Prehistory, University of Liège, Place du XX Août 7, 4000, Liège, Belgium

^e Shrieking Sixties, 1-3 Allée Lavoisier, 59650 Villeneuve-d'Ascq, France

^f University of Bordeaux, Proteome Platform, 146 rue Leo Saignat, F-33076 Bordeaux, France

^g Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paléo, Unité Evolution, Ecologie et Paléontologie, F-59000 Lille, France

*Corresponding author: fabrice.bray@univ-lille.fr

ORCID

Fabrice Bray: <https://orcid.org/0000-0002-4723-8206>

Stéphanie Flament: <https://orcid.org/0000-0002-0135-675X>

Dominique Bonjean : <https://orcid.org/0000-0002-0334-8307>

Grégory Abrams: <https://orcid.org/0000-0002-0616-4335>

Kevin Di Modica: <https://orcid.org/0000-0001-5067-5805>

Christian Rolando: <https://orcid.org/0000-0002-3266-8860>

Caroline Tokarski: <https://orcid.org/0000-0001-7296-3330>

Patrick Auguste: <https://orcid.org/0000-0002-8302-6307>

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Abstract

The ancient preserved molecules offer the opportunity to gain a better knowledge on the biological past. In recent years, bones proteomics has become an attractive method to study the animal biological origin, extinct species and species evolution as an alternative to DNA analysis which is limited by DNA amplification present in ancient samples and its contamination. However, the development of a proteomic workflow remains a challenge. The analysis of fossils must consume a low quantity of material to avoid damaging the samples. Another difficulty is the absence of genomic data for most of the extinct species. In this study, a proteomic methodology was applied to mammalian bones of 130,000 years old from the earlier Upper Pleistocene site of Scladina Cave (Belgium). Starting from 5 milligram samples, our results show a large majority of detected peptides matching collagen I alpha 1 and alpha 2 proteins with a sequence coverage up to 60%. Using sequence homology with modern sequences, a biological classification was successfully achieved and the associated taxonomic ranks to each bone were identified consistently with the information gained from osteomorphological studies and palaeoenvironmental and palaeodietary data. Among the taxa identified are the Felidae family, Bovinae subfamily, Elephantidae family and the *Ursus* genus. Amino acid substitutions on the collagens were identified providing new information on extinct species sequences and also helping in taxonomy-based clustering. Considering samples with no osteomorphological information, such as two bone retouchers, proteomics successfully identified the bovidae and ursidae families providing new information to the paleontologists on these objects. Combining osteomorphology studies and amino acid variations identified by proteomics, one retoucher was identified to be potentially from the *Ursus spelaeus* species.

Keywords: *Palaeoproteomics, Pleistocene, Scladina Cave, mammalian bones, taxonomy, peptides modifications, amino acids mutation*

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

1. Introduction

The study of bones by mass spectrometry (MS) has become an increasingly used method. Indeed, bones are real safes protecting their constitutive proteins. MS proved to be a robust and accurate method allowing the identification of proteins, their biological origins and their modifications (Cleland and Schroeter, 2018, Dallongeville, et al., 2016, Vinciguerra, et al., 2016, Welker, 2018) overpassing DNA analysis considering the longer temporal scales (Cappellini, et al., 2018, Demarchi, et al., 2016). For example, the identification of collagen I from *Brachylophosaurus canadensis* dinosaurs has been confirmed (Schroeter, et al., 2017) and collagen sequences from fossil *camels*, *Camelops* and c.f. *Paracamelus*, from the Arctic and sub-Arctic of Plio-Pleistocene North America has been reported (Buckley, et al., 2019). Another examples are the palaeoproteomics studies of *Gigantopithecus blacki* dated between the Early Pleistocene and the late Middle Pleistocene (Welker, et al., 2019) and Eurasian Rhinocerotidae of the Pleistocene epoch (Cappellini, et al., 2019).

Type I collagen is the major protein of bones. It is constituted by the association of two alpha 1 and one alpha 2 chains arranged in a super helix. The sequence of this protein is characterized by regular occurrences of glycine, proline and hydroxyproline (Shoulders and Raines, 2009). Non-collagenous proteins (NCPs) represent a smaller part of bone proteins (Young, et al., 1992). Type I collagen is the main protein identified by mass spectrometry in ancient bones.

Palaeoproteomic analysis of bones consists of a first protein/peptide preparation step, generally starting from several tens to hundreds milligrams of bones, and their analysis and identification by mass spectrometry and dedicated bioinformatics tools. The starting material amount is low in comparison to other techniques usually applied to bone analysis such as ¹⁴C dating (from several hundreds of milligrams to gram). The first step of a classical palaeoproteomic analysis is the demineralization which is performed most often by an acidic solution and seldom by a basic solution (Schroeter, et al., 2016). This step is well-established for decalcification and solubilization. The extracted proteins are then digested into peptides by an enzyme, most often the trypsin. Several digestion methods have been reported such as liquid digestion (Horn, et al., 2019, Sawafuji, et al., 2017), filter assisted sample preparation (FASP) (Cappellini, et al., 2014, Kostyukevich, et al., 2018) and solid digestion of demineralized bones (Cleland, 2018b). Recently was proposed a single-pot solid-phase-enhanced sample preparation (SP3) (Cleland, 2018a) which is an effective method to enable collagen and NCPs analyses along with the removal of coextracted humic compounds. Another method proposes to study the adhering collagen from the storage plastic bags (McGrath, et al., 2019). This technique allows studying the object without affecting the integrity of the artifact. Multi-protease approaches were also described to increase the protein sequence coverage (Lanigan, et al., 2020).

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Another recent studies in paleoproteomics also propose a digestion-free approach to characterize ancient remains (Cappellini, et al., 2019, Welker, et al., 2019).

Two techniques are used to identify the peptides extracted from bones: ZooMS and shotgun analysis based on LC separation hyphenated with ESI-MS. Developed in 2009, ZooMS (Buckley, et al., 2009) is based on peptide mass fingerprint analysis using MALDI-TOF MS. It allows taxonomic discrimination between a wide range of mammalian bones by the identification of peptides variants between species. ZooMS method has resulted in the identification of various species from ancient material such as hominin (Brown, et al., 2016), marine mammals (Hofman, et al., 2018), other mammals such as sheep, goat (Birch, et al., 2019, Brandt, et al., 2018), kangaroos (Buckley, et al., 2017a) from bones but also from derived material such as bones tools (Desmond, et al., 2018, Martisius, et al., 2020, McGrath, et al., 2019). Considering shotgun analysis, the amino acid sequences of the peptides are identified allowing thus an accurate identification of proteins including the species-related amino acids variants. The first sequencing experiments were proposed in the early 2000s on bone osteocalcin (Nielsen-Marsh, et al., 2002, Nielsen-Marsh, et al., 2005, Ostrom, et al., 2006, Ostrom, et al., 2000). Early 2010s showed the first application of shotgun analysis on a 43,000 year old woolly mammoth bone (Cappellini, et al., 2011) resulting in the identification of a hundred proteins. Since, the method has been extensively applied to the study of other extinct species such as *Castoroides ohioensis* (Cleland, et al., 2016), *Bison latifrons* (Hill, et al., 2015), ancient birds (Horn, et al., 2019), dinosaurs (Schroeter, et al., 2017). Along with the study of extinct species, phylogeny and evolutionary history are studied. Palaeoproteomics applied to the study of endemic South American 'ungulates' (Buckley, 2015), Darwin's South American ungulates (Welker, et al., 2015), ancient hominin specimens (Welker, et al., 2016), archaeological marine turtles (Harvey, et al., 2019) and tree sloth (Presslee, et al., 2019) are several of these applications.

We present here a robust proteomics workflow to study extinct species from Scladina Cave (Andenne, Belgium) dated from earlier Upper Pleistocene (130,000 years B.C.). The method was applied to 5 milligrams of starting material. Both ancient bones and retouchers were studied. Our analyses resulted in a successful characterization of biological classification and associated taxonomic ranks to each bone consistently with the information gained from morphological studies. Taxa identifications were also provided for the retouchers (not identified from their morphology) in agreement with the fauna found on the site. It can be pointed out that during this study, amino acid substitutions on the collagens were identified providing new information on extinct species sequences and also helping in taxonomy-based clustering.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

2. Materials and methods

2.1 Samples

Ten bones from the archaeological excavation of Scladina Cave were selected from different geological layers (Units 4B, 5, 6A, 6B, 6C, 4B). Scladina Cave is located 400 m southwest of the village of Sclayn located in Belgium between Andenne and Namur, close to the Meuse River. Several dating methods were established to estimate the site to be 130,000 years old (Otte, et al., 1983, Pirson, et al., 2008, Pirson, et al., 2014). The species of 8 bones on 10 were identified through their anatomy. The 2 remaining ones used as retouchers or “lisoirs” (smoothers) were from unknown or hypothetic origins. These two last bones are of particular interest because they are directly associated with retouchers by refitting. Bone. The Table 1 is referencing the 10 studied samples with archaeological information. Photographs of the studied bones are presented in Supplementary data file 1 Fig. S1.1 to S1.10.

Table 1

Bones samples from the Scladina Cave (Sclayn, Belgium).

Excavation year	Excavation number	Geological Unit	Location (square)	Description	Species	Sample name
1983	236-52	5	E13	Calcaneus	<i>Panthera spelaea</i>	Sc-1
2011	362-7	6C-LAGJ	G16	Metatarsal II	<i>Panthera spelaea</i>	Sc-2
1983	407-93	5	E12	Metatarsal III	<i>Panthera spelaea</i>	Sc-3
1981	127-3	6A	C4	Humerus	<i>Panthera pardus</i>	Sc-4
2013	46-2	6B	C37	Metatarsal III	<i>Ursus spelaeus</i>	Sc-5
2013	44-8	6A	D37	Metatarsal I	<i>Ursus thibetanus</i>	Sc-6
2003	590-2	6A	B36	Metatarsal III	<i>Ursus thibetanus</i>	Sc-7
1995	430-97	4B	C36	Rib	<i>Mammuthus</i> sp	Sc-8
1987	52	5	I22	Retoucher	Unknown	Sc-9
1985	F16-17	5	F16	Retoucher	Unknown	Sc-10

2.2 Chemicals and biochemicals.

All aqueous solutions were prepared from ultrapure grade water obtained by water filtration with a two stages Millipore system (Milli-Q® Academic with a cartouches Q-Gard 1 and Progard 2, Merck Millipore, Burlington, Massachusetts, United States). All chemicals, biochemicals and solvents were purchased from Sigma-Aldrich (Saint-Louis, Missouri, USA) and used without purification. All solvents were MS analytical grade.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

2.3 Protein extraction from bones and delipidation.

Bone fragments were mechanically ground in fine powder with a pestle in an agate mortar. 5 mg of bone powder was transferred to a 2 mL Eppendorf® tube (Eppendorf; Westbury, NY, USA) and 1 mL of demineralizing solution (5% trifluoroacetic acid (TFA) in water) was added. The solution was stirred during 24 h at 4 °C and then centrifuged for 10 min at 10,000 g at room temperature. The bone powder and the demineralizing solution were carefully separated. 1 mL of pure water was added to the centrifuged bone powder and the solution was stirred for 10 min at 4 °C. The solution was centrifuged for 5 min at 10,000 g, then the supernatant was removed. This step was repeated a second time. The delipidation was performed by adding 100 µL of water to the bone powder, then 900 µL of a chloroform / methanol solution (2/1 v/v). The sample was stirred for 1 h at 4 °C and then centrifuged for 10 min at 10,000 g. The lower phase that contains the lipids was removed. The bone powder was washed again twice with water. Residual water in bone powder was then evaporated at room temperature with a SpeedVac™ Concentrator (Eppendorf™ Concentrator Plus, Eppendorf, Hamburg, Germany). The recovered demineralizing solution was evaporated at room temperature with a SpeedVac™. Then 200 µL of lysis buffer (8 M urea, 4 % SDS, 0.2% DCA, 50 mM DTT, 100 mM ammonium bicarbonate pH 8.8) were added to the bone powder and on the evaporated demineralisation solution. The suspensions were incubated overnight with shaking at 4 °C, before the eFASP digestion step.

2.4 eFASP digestion

Before use, 0.5 mL Amicon® ultra centrifugal filters with a cut-off of 10 kDa (EMD Millipore, Darmstadt, Germany) were freshly incubated overnight with the passivation solution containing 5% (v/v) Tween®-20. Amicon® ultra centrifugal filters filled with 200 µL water were poured in a water bath for 20 min each four times. The bone powder suspension and the demineralized fraction were transferred to separate Amicon® filters, 100 µL of exchange buffer was added (8 M urea, 0.2% DCA (deoxycholic acid) w/w, 100 mM ammonium bicarbonate pH 8.8) and then Amicon® filters centrifuged for 30 min at 10,000 g. The filtrates were discarded. 200 µL microliters of exchange buffer were added again into the Amicon® filters which were centrifuged. This step was repeated twice. The proteins were alkylated during 1 h at room temperature in the dark using 100 µL of alkylation buffer (8 M urea, 50 mM iodoacetamide, and 100 mM ammonium bicarbonate, pH 8.8). The Amicon® filters were centrifuged for 30 min at 10,000 g and the filtrate was discarded. After the alkylation step, 200 µL of exchange buffer was added to the Amicon® filter which was centrifuged for 30 min at 10,000 g and the filtrate was discarded. Two hundred microliters of digestion buffer (0.2% DCA (deoxycholic acid) w/w, 50 mM ammonium bicarbonate pH 8.8) were added to the Amicon® filter and then centrifuged. This step was repeated twice, discarding the filtrate at each step. The Amicon® filters were transferred

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

to new 2 mL tubes. 100 μ L microliters of digestion buffer and 40 μ L of trypsin / LysC (Promega, Madison, USA) were added and incubated with shaking in a heating block tube (MHR23, Hettich, Netherlands) overnight at 37 °C. After this step, the peptides present in the Amicon® filter were recovered in the lower tube by centrifugation for 15 min at 10,000 g. In order to obtain a maximum of peptides, the Amicon® filters were washed twice with 50 μ L of ammonium bicarbonate solution (50 mM pH 8.8). The filtrates containing all the peptides were transferred to 1.5 mL Eppendorf® tubes. 200 μ L microliters of ethyl acetate and 2.5 μ L of TFA were added inducing the appearance of a white precipitate. At once, 800 μ L of ethyl acetate were added again and the Eppendorf® tubes were centrifuged for 10 min at 10,000 g. The organic phase was eliminated. This step was repeated twice. The Eppendorf® tube were placed for 5 min at 60 °C in a heating block tubes (SBH130, Stuart, Staffordshire, UK) to remove the remaining ethyl acetate. The samples were evaporated to dryness at room temperature with a SpeedVac™ Concentrator. The wall of the Eppendorf® tubes were rinsed with 100 μ L of a methanol / water (50/50 v/v) mixture and again evaporated to dryness. For mass spectrometry analysis, the sample was dissolved again in 10 μ L of water solution containing 0.1% of formic acid. The concentration was then estimated by measuring the OD at 215 nm using 1 μ L of the solution using a droplet UV spectrometer (DS-11+, Denovix, Wilmington, USA). The samples were diluted at a concentration of 1 μ g/ μ L before analysis.

2.5 *Liquid chromatography-tandem mass spectrometry*

LC-MS/MS analyses were performed on an Orbitrap Q Exactive plus mass spectrometer hyphenated to a U3000 RSLC Microfluidic HPLC System (ThermoFisher Scientific, Waltham, Massachusetts, USA). 1 μ L of the peptide mixture at a concentration of 1 μ g/ μ L was injected with solvent A (5% acetonitrile and 0.1% formic acid v/v) for 3 min at a flow rate of 10 μ L.min⁻¹ on an Acclaim PepMap100 C18 pre-column (5 μ m, 300 μ m i.d. \times 5 mm) from ThermoFisher Scientific. The peptides were then separated on a C18 Acclaim PepMap100 C18 reversed phase column (3 μ m, 75 mm i.d. \times 500 mm), using a linear gradient (5-40%) of solution B (75% acetonitrile and 0.1% formic acid) at a rate of 250 nL.min⁻¹ in 160 minutes and then 100% of solution B in 5 minutes. The column was washed for 5 minutes with buffer B and then re-equilibrated with buffer A. The column and the pre-column were placed in an oven at a temperature of 45 °C. The total duration of the analysis was 180 min. The LC runs were acquired in positive ion mode with MS scans from m/z 350 to 1,500 in the Orbitrap mass analyser at 70,000 resolution at m/z 400. The automatic gain control was set at 1e106. Sequentially MS/MS scans were acquired in the high-energy collision dissociation cell for the 15 most-intense ions detected in the full MS survey scan. Automatic gain control was set at 5e105, and the normalized collision energy was set to 28 eV. Dynamic exclusion was set at 90 s and ions with 1 and more than 8 charges were excluded.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

2.6 Data analysis

Proteomics data were processed with Mascot (version 2.5.1, Matrix Science, London, UK) against NCBI database (NCBIProt_20190815) restricted to Mammalian (5,242,539 sequences). Error tolerant searches with up to three missed cleavages, 10 ppm mass error for MS and 15 mmu for MS/MS. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation and asparagine, glutamine deamidation were selected as variable modifications. The automatic error tolerant search which search the modified forms of already identified peptides was allowed. The second bioinformatics analysis was performed with PEAKS X (Bioinformatics software, Waterloo, Canada) against a home-made database containing 1,765 collagen sequences extracted from NCBI database (All_Collagen, downloaded 08-2019) restricted to Mammalian. Precursor's mass tolerance was fixed to 10 ppm and fragment ion mass tolerance to 0.02 Da. Three missed cleavages were allowed. The same post-translational modifications (PTMs) above were allowed plus hydroxylation of amino acids (RYFPNKD) as variable modifications. Five variable PTMs were allowed per peptides. PEAKS PTM and SPIDER ran with the same parameters. Results were filtered using the following criteria: protein score $-10\log P \geq 20$, 1% peptide False Discovery Rate (FDR), PTM with Ascore = 20, mutation ion intensity = 5% and *Denovo* ALC $\geq 50\%$. Peptides with amino acids substitutions was filtered with minimal intensity set as $1E+7$. Peptides identified on collagen I alpha 1 and I alpha 2 were aligned against the NCBI non redundant protein sequence (all non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects containing 308,570,119 sequences) to find similarity with BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). The scoring parameter alignment used BLOSUM62 matrix. The specific peptides were validated with a score of 100% of identity and full query coverage. Fragment ion spectra for specific peptides, deamidated and oxidative peptides and novel amino acid substitution biomarkers were manually examined for quality (Supplementary data 3, 5 and 6). The mass spectrometry proteomics data have been deposited on the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (Vizcaíno, et al., 2012) with the data set identifier PXD021171.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

3. Results

3.1 Bone samples

The bones and artefacts studied in this paper were recovered in Scladina Cave located 400 m southwest of the village of Sclayn (Andenne, Namur, Belgium), close to the Meuse River (Fig. 1). The site has been under scientific excavation since 1978. The cave infilling indicates a sequence covering at least the end of the Middle Pleistocene up to the Holocene (Pirson, et al., 2008, Pirson, et al., 2014). The cave stratigraphy indicates a sequence of the Upper Pleistocene (Pirson, et al., 2008). Lithic artefacts and faunal materials are found through the entire sedimentary sequence (Abrams, et al., 2010, Bonjean, et al., 2009). The fauna is composed of herbivorous species including horse (*Equus ferus*), red deer (*Cervus elaphus*), fallow deer (*Dama dama*), chamois (*Rupicapra rupicapra*), aurochs or bison (*Bos primigenius* or *Bison priscus*), woolly rhinoceros (*Coelodonta antiquitatis*). Omnivorous species including cave bear (*Ursus spelaeus*), brown bear (*Ursus arctos*), red fox (*Vulpes vulpes*) and arctic fox (*Alopex lagopus*), as well as carnivorous species including cave lion (*Panthera spelaea*), panther (*Panthera pardus*), cave hyaena (*Crocuta crocuta spelaea*) and wolf (*Canis lupus*). In the cave modified bones with cutting marks, filleting marks or breakage damages were also recovered (Otte, et al., 2000).

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

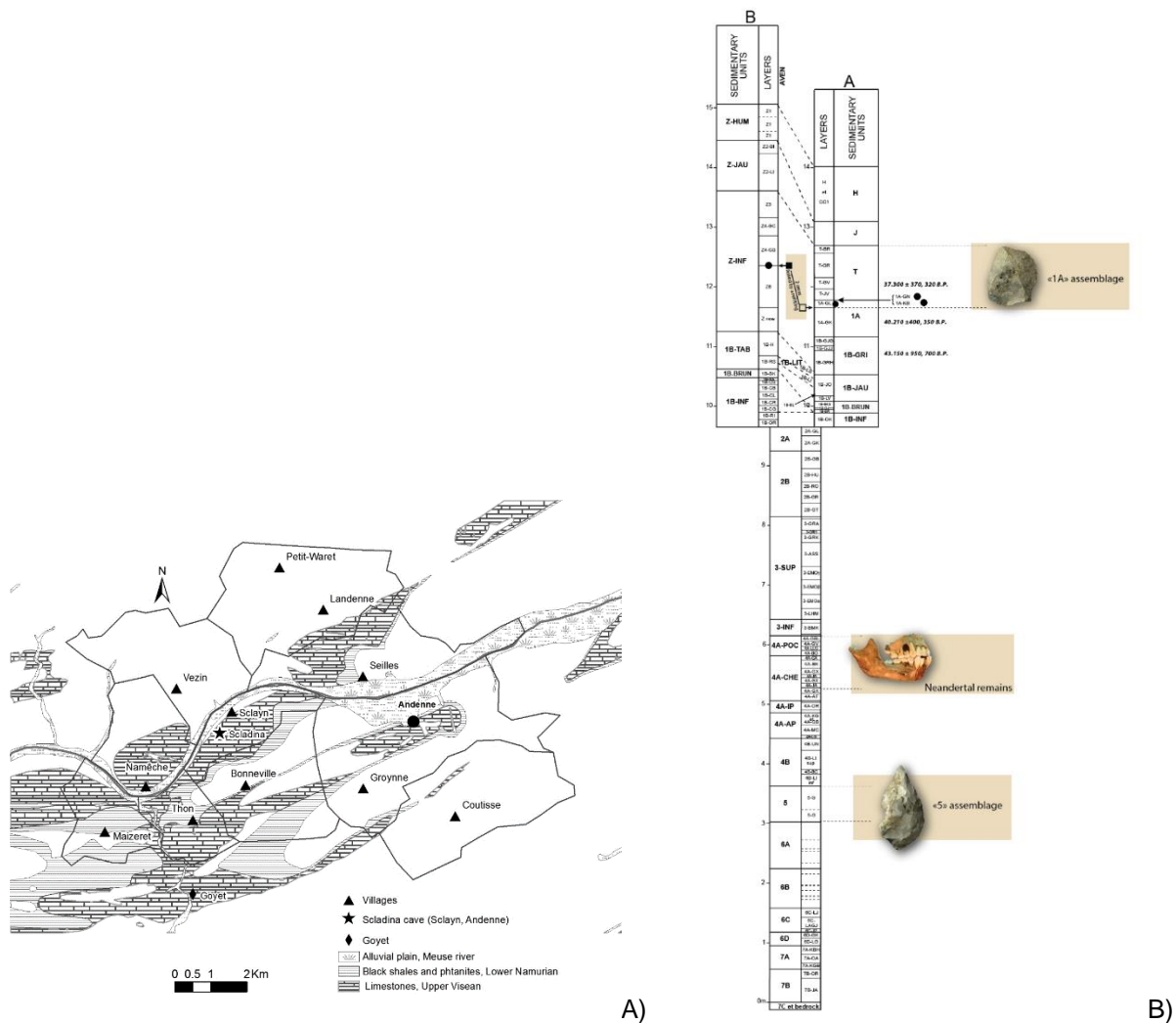


Fig. 1. A) Location of Scladina Cave. Grey areas: distribution of the Palaeozoic carbonated rocks in Belgium. B) Stratigraphic profile of Scladina Cave composed of the entrance-sequence (lower and upper portions of A) and the aven sequence (upper portion of B; adapted from Bonjean et al 2015) (Bonjean, et al., 2015).

3.2 Proteomics methodology and identified proteins

We applied LC-MS/MS shotgun proteomics to ten selected 130,000 year old bones (Supplementary data 1). The method was applied to 5 mg starting material resulting in two analyzed fractions, the bone powder fraction and the demineralization residual fraction. Both fractions were digested with the eFASP method to increase the sensitivity, recovery and coverage of the identified proteins, and decrease the sensitivity to chemical contaminations (Erde, et al., 2014). As the trypsin enzyme is able to digest directly the proteins contained in bone powder (Schroeter, et al., 2019) no protein extraction was carried out on both fractions and trypsin was reacted directly on the bone powder and demineralization residual fractions. For bone powder fractions the mean Mascot score for identified peptides for the best hit score is 639 and 630 for demineralization residual fractions. The coverage sequence is roughly similar for both fractions 58 and 57 for the bone powder and demineralization residual fractions (Table 2). The Mascot score is very high for all identified proteins

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

which tends to show a very good conservation of proteins inside the bones. For bone powder fractions the mean Mascot protein best hit score is 1618 and 1403 for demineralization residual fractions. The difference between the Mascot scores for both fractions is significantly different showing that more proteins are present in the bone powder fraction as expected as the demineralization step decalcifies the bone sample and remove the contaminants as keratins (Schroeter, et al., 2016). In our work, we used an aqueous 5% TFA demineralizing solution. TFA is a solvent commonly used in reverse-phase HPLC protein purification because of its effectiveness in solubilizing hydrophobic peptides. Using TFA, proteins or peptides are less modified as the TFA pKa is lower than HCl pKa which is the acid commonly used in paleoproteomics for demineralization (0.3 and -6.3 respectively). For all samples and fractions (i.e. bone powder and demineralization residual), the majority of the identified peptides matches mammalian collagen I alpha 1 (COL1A1) and collagen I alpha 2 (COL1A2) (Table 2). Considering COL1A1 and COL1A2 proteins, the peptides are distributed all along the sequences. The sequence coverage average for the best hit is 57% excluding the signal peptide and N, C-terminus propeptides. The other identified proteins with a low number of peptides are type II cytoskeletal 1 keratin, hornerin (less than 20 peptides for both proteins) which are contaminants of archaeological excavations (Hendy, et al., 2018). Others collagens are identified in both fractions for all samples as collagen II alpha 1 (COL2A1), collagen III alpha 1 (COL3A1), collagen V alpha 2 (COL5A2) and fibrillary collagen NC1 domain (Supplementary data 1).

Table 2

Information related to the quality of MS analysis for demineralization residual fraction (Dem.) and bone powder fraction (Pow.), including number of peptide matches above identity and homology thresholds, number of identified proteins, best hit and its NCBI accession number, Mascot protein score, best hit protein percentage of coverage. Family and genus taxa identified in archaeological samples (see sections 3.3 and 3.6). All identified proteins are listed in Supplementary data 2.

Sample	Fraction	Peptides: identity	Peptides: homology	Identified proteins	Best Hits Protein name [<i>species</i>]	Accession	Score	Coverage (%)	Family	Genus
Sc-1	Dem.	636	839	99	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,269	66	Felidae	<i>Panthera</i>
	Pow.	656	847	98	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,673	66	Felidae	<i>Panthera</i>
Sc-2	Dem.	561	731	89	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,357	61	Felidae	<i>Panthera</i>
	Pow.	623	853	106	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,625	63	Felidae	<i>Panthera</i>
Sc-3	Dem.	570	746	90	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,420	63	Felidae	<i>Panthera</i>
	Pow.	543	713	85	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,692	57	Felidae	<i>Panthera</i>
Sc-4	Dem.	577	732	76	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,488	66	Felidae	<i>Panthera</i>
	Pow.	531	717	79	COL1A1 isoform X3 [<i>Panthera pardus</i>]	XP_01927135 1.1	1,635	63	Felidae	<i>Panthera</i>
Sc-5	Dem.	657	836	98	COL1A1 isoform X1 [<i>Ursus arctos horribilis</i>]	XP_02636891 3.1	1,642	53	Ursidae	<i>Ursus</i>
	Pow.	721	924	95	COL1A1 [<i>Ailuropoda melanoleuca</i>]	XP_00292377 5.1	1,635	55	Ursidae	<i>Ailuropoda</i>

*Extinct species identification from Upper Pleistocene bone fragments
not identifiable from their osteomorphological studies by proteomics analysis*

Sc-6	Dem.	660	837	88	COL1A1 [<i>Ailuropoda melanoleuca</i>]	XP_00292377 5.1	1,793	55	Ursidae	<i>Ailuropoda</i>
	Pow.	709	915	97	COL1A1 [<i>Ailuropoda melanoleuca</i>]	XP_00292377 5.1	2,120	56	Ursidae	<i>Ailuropoda</i>
Sc-7	Dem.	677	860	111	COL1A1 isoform X1 [<i>Ursus arctos horribilis</i>]	XP_02636891 3.1	1,447	54	Ursidae	<i>Ursus</i>
	Pow.	701	950	104	COL1A1 isoform X1 [<i>Ursus arctos horribilis</i>]	XP_02636891 3.1	1,689	52	Ursidae	<i>Ursus</i>
Sc-8	Dem.	855	1156	134	COL1A2 [<i>Mammut americanum</i>]	P85154.3	1,246	61	Mammutidae	<i>Mammut</i>
	Pow.	561	750	78	COL1A1 [<i>Loxodonta africana</i>]	XP_01059264 4.1	1117	52	Elephantidae	<i>Loxodonta</i>
Sc-9	Dem.	532	715	89	COL1A1 [<i>Bubalus bubalis</i>]	XP_00604121 4.2	1149	53	Bovidae	<i>Bubalus</i>
	Pow.	728	943	98	COL1A1 [<i>Bubalus bubalis</i>]	XP_00604121 4.2	1607	53	Bovidae	<i>Bubalus</i>
Sc-10	Dem.	577	748	83	COL1A1 [<i>Ailuropoda melanoleuca</i>]	XP_00292377 5.1	1224	53	Ursidae	<i>Ailuropoda</i>
	Pow.	620	771	79	COL1A1 isoform X1 [<i>Ursus arctos horribilis</i>]	XP_02636891 3.1	1389	53	Ursidae	<i>Ursus</i>

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

3.3 Taxa classification.

All analyses were carried out blind, especially for taxa classification. The identification of taxa discriminant peptides of COL1A1 and COL1A2 was performed using BLASTp analysis against the non-redundant NCBI database. In particular, we used the peptides from the taxa identified by LC-MS/MS analyses. We used for (i) *Panthera* taxa, peptides identified from *Panthera tigris altaica*, *Panthera pardus*, *Felis silvestris catus*, *Acinonyx jubatus*, *Puma concolor* and *Lynx pardinus*; (ii) for *Ursus* taxa, peptides identified from *Ursus maritimus*, *Ursus arctos horribilis*, *Ailuropoda melanoleuca*, (iii) for *Loxodonta* taxa peptides from *Loxodonta Africana*; *Mammut americanum* and (iv) for the *Bos* taxa peptides from *Bos primigenius taurus*, *Bos grunniens mutus*, *Bos primigenius indicus*, *Bubalus Bubalis*, *Bison bison*.

Table 3 shows the unique peptides identified in the studied samples. Their corresponding MS/MS spectra are shown in Supplementary Data 3 (Fig. S3.1 – S3.19). All these unique peptides were identified in both fractions, i.e. demineralization residual and bone powder. Most of the identified peptides carry post-translational modifications like proline/methionine oxidation, glutamine/asparagine deamidation and carboxymethyl arginine. These modifications associated with protein biological ageing are found in analyzes of ancient bones (Cappellini, et al., 2011, Schroeter and Cleland, 2016, Wadsworth and Buckley, 2014).

Discriminant peptides matching the Ursidae family were identified in the samples Sc-5, Sc-6, Sc-7.

Considering the sample Sc-8, the unique peptides match the Proboscidea order and 2 unique peptides (RGPNGEAGSAGPAGPPGLR and GPNGEAGSAGPAGPPGLR) show specificity to *Mammut americanum* species. The Proboscidea order taxa classification corresponds to the morphological identification realized by palaeontologists.

Considering the Sc-9 and Sc-10 retouchers which were not identified by palaeontologist from the bone appearance, discriminant peptides for the Bovinae subfamily matching the *Bos primigenius taurus*, *Bos mutus*, *Bos indicus*, *Bubalus Bubalis*, *Bison bison* species as well as discriminant peptides for Ursidae family were identified respectively in samples Sc-9 and Sc-10. It can be remarked that 2 unique peptides from the sample Sc-10 were found in the samples Sc-5, Sc-6 and Sc-7.

Table 3

Unique peptides identified in the studied archaeological samples by LC-MS/MS. Underlined amino acids carry posttranslational modifications. The list of all the identified peptides and their modifications is given in Supplementary data 4.

Sample	Peptide sequence	Modifications	Classification
Sc-1	GPSGAVGAP <u>GV</u> <u>N</u> GAP <u>GE</u> AGR	2 oxidations (P), deamidation (N)	Family: <i>Felidae</i>
	GPAGPSG <u>P</u> <u>M</u> G <u>K</u> DGR	Oxidation (M), carboximethyl (R)	Family: <i>Felidae</i>
Sc-2	GPSGAVGAP <u>GV</u> <u>N</u> GAP <u>GE</u> AGR	2 oxidations (P), deamidation (N)	Family: <i>Felidae</i>
	GPAGPSG <u>P</u> <u>M</u> G <u>K</u> DGR	Oxidation (M), carboximethyl (R)	Family: <i>Felidae</i>
Sc-3	GPSGAVGAP <u>GV</u> <u>N</u> GAP <u>GE</u> AGR <u>D</u> G <u>N</u> <u>P</u> G <u>N</u> D <u>G</u> P <u>P</u> G <u>R</u>	4 oxidations (P), 3 deamidations (N)	Family: <i>Felidae</i>
	GPSGAVGAP <u>GV</u> <u>N</u> GAP <u>GE</u> AGR	2 oxidations (P), deamidation (N)	Family: <i>Felidae</i>
Sc-4	GPSGAVGAP <u>GV</u> <u>N</u> GAP <u>GE</u> AGR	2 oxidations (P), deamidation (N)	Family: <i>Felidae</i>
	GPAGPSG <u>P</u> <u>M</u> G <u>K</u> DGR	Oxidation (M), carboximethyl (R)	Family: <i>Felidae</i>
Sc-5	GESG <u>N</u> K <u>G</u> EP <u>G</u> SVG <u>P</u> Q <u>G</u> PP <u>G</u> PSGEEGK	2 oxidations (P)	Family: <i>Ursidae</i>
	GE <u>P</u> GSV <u>G</u> P <u>Q</u> G <u>P</u> PP <u>G</u> PSGEEGK	2 oxidations (P), deamidation (Q)	Family: <i>Ursidae</i>
	GEPGPAGAVGPV <u>G</u> AVGPR	No modification	Family: <i>Ursidae</i>
	HG <u>N</u> RGEPGPAGAVGPV <u>G</u> AVGPR	Deamidation (N)	Family: <i>Ursidae</i>
	GEPGPAGAVGPV <u>G</u> AVGPR	No modification	Family: <i>Ursidae</i>
	AVGPV <u>G</u> AVGPR	No modification	Family: <i>Ursidae</i>
Sc-6	GESG <u>N</u> K <u>G</u> EP <u>G</u> SVG <u>P</u> Q <u>G</u> PP <u>G</u> PSGEEGK	2 oxidations (P)	Family: <i>Ursidae</i>
	GE <u>P</u> GSV <u>G</u> P <u>Q</u> G <u>P</u> PP <u>G</u> PSGEEGK	2 oxidations (P), deamidation (Q)	Family: <i>Ursidae</i>
	GEPGPAGAVGPV <u>G</u> AVGPR	No modification	Family: <i>Ursidae</i>
	GETGLRGD <u>V</u> GNPGR	No modification	Family: <i>Ursidae</i>

*Extinct species identification from Upper Pleistocene bone fragments
not identifiable from their osteomorphological studies by proteomics analysis*

	HG <u>N</u> RGEPGPAGAVGPVGA <u>V</u> GPR	Deamidation (N)	Family: <i>Ursidae</i>
Sc-7	GESG <u>N</u> KGEPGSVGPQGP <u>P</u> GPSGEEGK	2 oxidations (P)	Family: <i>Ursidae</i>
	GE <u>P</u> GSVGPQGP <u>P</u> GPSGEEGK	2 oxidations (P), deamidation (Q)	Family: <i>Ursidae</i>
	HG <u>N</u> RGEPGPAGAVGPVGA <u>V</u> GPR	Deamidation (N)	Family: <i>Ursidae</i>
Sc-8	GYPGNAGPVGTAGAPG <u>P</u> QGPLG <u>P</u> AGK	Deamidation (N)	Order: <i>Proboscidea</i>
	<u>Q</u> GPSGSSGERG <u>P</u> GPAGPPGLAGPPGESGR	Oxidation (P), deamidation (Q)	Order: <i>Proboscidea</i>
	TG <u>P</u> PG <u>P</u> AGITG <u>P</u> PG <u>P</u> GAAGK	5 oxidations (P)	Order: <i>Proboscidea</i>
	GEAGL <u>P</u> GVSGPVGP <u>P</u> GN <u>P</u> GANGLAGAK	3 oxidations (P), 2 deamidations (N)	Order: <i>Proboscidea</i>
	RGP <u>N</u> GEAGSAG <u>P</u> AG <u>P</u> GLR	2 oxidations (P), deamidation (N)	Species: <i>Mammut americanum</i>
	GP <u>N</u> GEAGSAG <u>P</u> AG <u>P</u> GLR	2 oxidations (P)	Species: <i>Mammut americanum</i>
	GESGSKGEPGSAGPQGP <u>P</u> GPSGEEGK	2 oxidations (P)	Order: <i>Proboscidea</i>
Sc-9	GEGGP <u>Q</u> G <u>P</u> RGSEGPQGV <u>R</u>	2 deamidation (Q, R)	Subfamily: <i>Bovinae</i>
	GAPGAIGAPGPAGAN <u>G</u> DRGEAGPAGPAGPAGPR	2 oxidations (P), deamidation (N)	Subfamily: <i>Bovinae</i>
	GYPGNAGPVGAAGAPGPQGPVGPV <u>G</u> K	2 oxidations (P)	Subfamily: <i>Bovinae</i>
Sc-10	GESG <u>N</u> KGEPGSVGPQGP <u>P</u> GPSGEEGK	2 oxidations (P)	Family: <i>Ursidae</i>
	GE <u>P</u> GSVGPQGP <u>P</u> GPSGEEGK	2 oxidations (P), deamidation (Q)	Family: <i>Ursidae</i>

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Considering the Sc-1, Sc-2, Sc-3 and Sc-4 samples, discriminant peptides for the Felidae family were identified (Table 3) however, the identified peptides match all the following species (except *Lynx pardinus*): *Panthera tigris altaica*, *Panthera pardus*, *Felis silvestris catus*, *Acinonyx jubatus* and *Puma concolor*. Samples Sc-1 to Sc-3 were identified by paleontologists as *Panthera spelaea* bones and Sc-4 as *Panthera pardus* bone. The phylogenetic data shows that their closest ancestors are *Panthera leo* (Barnett, et al., 2016). Proteomics results may be refined allowing to trace amino acid mutations by introducing phylogenetic information. The species of closest ancestors were used to screen amino acid mutations. In our case only modern *Panthera pardus* collagen sequences are present in the databases and not *Panthera leo* sequences. So for samples Sc-1 to 4, collagen amino acids variations has been be studied from modern *Panthera pardus* sequences. Table 4 shows the closest species in the databases for the 10 studied bones and the phylogenetic data for each species. For retouchers, the faunal data found on the site, information from paleontologists, the results of proteomic analyses (number of peptides, sequence coverages) and phylogenetic data allowed targeting the closest species.

Table 4

Identification of the studied bones taxa by paleontologists and proteomic analysis and their closest species in the public proteins databases.

Sample	Identified taxa		Closest species in databases	Phylogenetic reference
	Anatomy ^a	Proteomics ^b		
Sc-1	<i>Panthera spelaea</i>	Felidae	<i>Panthera pardus</i>	(Barnett, et al., 2016)
Sc-2	<i>Panthera spelaea</i>	Felidae	<i>Panthera pardus</i>	
Sc-3	<i>Panthera spelaea</i>	Felidae	<i>Panthera pardus</i>	
Sc-4	<i>Panthera pardus</i>	Felidae	<i>Panthera pardus</i>	
Sc-5	<i>Ursus spelaeus</i>	Ursus	<i>Ursus arctos horribilis</i>	(Veitschegger, et al., 2018)
Sc-6	<i>Ursus thibetanus</i>	Ursus	<i>Ursus arctos horribilis</i>	(Wu, et al., 2015)
Sc-7	<i>Ursus thibetanus</i>	Ursus	<i>Ursus arctos horribilis</i>	
Sc-8	<i>Mammuthus sp</i>	Proboscidea	<i>Ixodonta africana</i>	(Rohland, et al., 2007)
Sc-9	Unknown	Bovinae	<i>Bos primigenius</i> <i>Taurus</i>	(Park, et al., 2015)
Sc-10	Unknown	Ursus	<i>Ursus arctos horribilis</i>	(Veitschegger, et al., 2018)

^aSpecies. ^bOrder, family, subfamily or genus.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

3.4 Asparagine and glutamine deamidation.

The analysis of posttranslational modifications such as asparagine and glutamine deamidation or proline oxidation informs on the state of preservation of bones. Ancient, or at least more damaged, samples have a higher level of deamidation (Hill, et al., 2015, Ramsøe, et al., 2020, Schroeter and Cleland, 2016, Van Doorn, et al., 2012, Welker, et al., 2015). To calculate the frequency of deamidation and oxidation in all samples, 2 distinct peptides (GQAGVMGFPGPK and GANGAPGIAGAPGFPGAR) common to all the identified sequences of COL1A1 were used (Supplementary data 5, table S1). The quantified peptide modifications are oxidation on methionine or proline (m/z increment: + 15.9949), deamidation on glutamine or asparagine (m/z increment: + 0.9840). The sequence, mass, retention and fragmentation spectra of the native and modified peptides are given in Supplementary data 5 (Fig. S5.1 – S5.6).

The frequency of modification is calculated according to the formula (1):

$$\text{Frequency of modification} = \frac{\text{All Intensity of modified peptide}}{\text{All Intensity of modified peptide} + \text{intensity of control peptide}} \times 100 \quad (1)$$

For COL1A1, in the bone powder fraction the average percentage of deamidation for all samples is 76.4% for glutamine and 85.5% for asparagine. These percentages are slightly higher in the demineralization residual fraction reaching 83.4% for glutamine and 91.4% for asparagine. The percentage of oxidation of prolines is stable at 99.1% for powder fractions and 99.4% for demineralization residual fraction (Supplementary data 5, Fig. S5.13). The same results are observed on COL1A2 (Supplementary data 5, Fig. S5.14). For COL1A1, in the bone powder fraction the average percentage of deamidation for all samples is 67.1% for glutamine and 80.7% for asparagine. These percentages are slightly higher in the demineralization residual fraction reaching 83% for glutamine and 80.9% for asparagine. The percentage of oxidation of prolines is stable at 99.6% for powder fractions and 99.7% for demineralization residual fraction (Supplementary data 5, Fig. S14). All samples exhibits elevated deamidation frequencies in the range expected from Middle Pleistocene bones (Lanigan, et al., 2020, Welker, et al., 2017). Pearson correlations show that the best correlation between bone powder fraction and the demineralization residual fraction is asparagine deamidation ($r = 0.78$ for COL1A1 and $r = 0.62$ for COL1A2) which proves that it is less impacted by the demineralization procedure.

3.6 Amino acids variation

Amino acid variants not referenced in the protein databases provide potential phylogenetic information. These variants may be found by error-tolerant search of MS/MS spectra, which is a computer intensive data processing. The very good conservation of proteins in bones from the

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Scladina Cave helped this study favorably. Supplementary data 6 (Fig. S6.1 – S6.56) shows the MS/MS spectra of peptides with amino acid substitutions for all samples and table 5 indicates the peptides with amino acids mutations. The analysis of variations of amino acids was carried out on the protein sequences of the species closest to the species identified by the paleontologists (Table 3).

Table 5

Peptides and amino acids variations identified in the studied archaeological samples. For each sample, the following information is described: peptides sequences, mass over charges (m/z), charge of peptides (z), the position and the type of mutation, the collagen chain, accession number and species. The peptides sequences are referencing the identified modifications (oxidation, deamidation, substitution). A m/z increase of +15.9949 corresponds to the oxidation of proline m/z increase + 0.9840 to deamidation of asparagine (N) or glutamine (Q). (sub *) denotes the substitution of amino acid where * is amino acid.

Samples	Peptide	m/z	Z	Position of mutation	Mutation	Protein	Accession number	Species
Sc-1 Sc-2 Sc-3	N(sub G)EAGPSGPAGPTGAR	669.8234	2	779	G → N	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	GEQ(sub A)GPSGPAGPTGAR	669.8234	2	781	A → Q	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	GEAGPSGPAGPTGAE(sub R)GAP(+15.99)GDR	912.4112	2	793	R → E	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	N(sub G)PAGPAGPIGPVGAR	665.8649	2	1067	G → N	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	GPTGDP(+15.99)GKN(+.98)GDKGD(sub H)AGLAGAR	482.7312	4	513	H → D	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
	TN(sub G)ETGASGPP(+15.99)GFAGEK	768.3517	2	833	G → N	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
	TGEG(sub T)GASGPP(+15.99)GFAGEK	717.8282	2	835	T → G	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
	N(sub G)PSGAVGAP(+15.99)GVNGAP(+15.99)GEAGR	884.4193	2	908	G → N	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
	D(sub H)GN(+.98)RGEP(+15.99)GPAGVVPVAVGPR	678.3462	3	976	H → G	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
	D(sub G)N(+15.99)RGEP(+15.99)GPAGVVPVAVGPR	996.0010	2	977	G → D	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
N(sub G)PTGPQGIR	470.2547	2	998	G → N	COL1A2	XP_019298882.1	<i>Panthera pardus</i>	
TGD(sub H)PGSVGPAGVR	585.2981	2	1071	H → D	COL1A2	XP_019298882.1	<i>Panthera pardus</i>	
Sc-4	GDAGPD(sub P)GPAGPTGPP(+15.99)GPIGN(+.98)VGAP(+15.99)GPK	1171.059	2	838	P → D	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	N(+.98)GDE(sub R)GETGPAGPAGPIGPVGAR	989.4691	2	1063	R → E	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	GEAGPSGPAGPTGAE(sub R)GAP(+15.99)GDR	912.4112	2	793	R → E	COL1A1	XP_019271349.1	<i>Panthera pardus</i>
	GPTGDP(+15.99)GKN(+.98)GDKGD(sub H)AGLAGAR	482.7312	4	513	H → D	COL1A2	XP_019298882.1	<i>Panthera pardus</i>
Sc-5	N(sub G)DAGPP(+15.99)GPAGPTGPP(+15.99)GPIGN(+.98)VGAPGPK	794.3891	3	833	G → N	COL1A1	XP_026368914.1	<i>Ursus arctos horribilis</i>
	AGEE(sub K)GHAGLAGAR	399.1977	3	511	K → E	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	AGEKGD(sub H)AGLAGAR	586.8037	2	513	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(sub G)ETGAHGPP(+15.99)GFAGEK	495.5627	3	833	G → N	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>

*Extinct species identification from Upper Pleistocene bone fragments
not identifiable from their osteomorphological studies by proteomics analysis*

	TGETGQ(sub A)HGPP(+15.99)GFAGEK	529.2446	3	837	A → Q	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(sub G)PP(+15.99)GAVGAP(+15.99)GVNGAP(+15.99)GEAGR	896.9307	2	908	G → N	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(+.98)IGPVGAVGAP(+15.99)GPD(sub H)GPVGP TG	644.3332	3	967	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	GN(+.98)IGPVGAVGAP(+15.99)GPD(sub H)GPVGP TGK	689.3215	3	967	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	D(sub H)GN(+.98)RGEP(+15.99)GPAGAVGPVAVGPR	668.6663	3	976	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	D(sub G)N(+15.99)RGEP(+15.99)GPAGAVGPVAVGPR	981.4858	2	977	G → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	GPSGPQ(+.98)GVRGDE(sub K)GEP(+15.99)GDKGPR	689.3215	3	1009	K → E	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	IGG(sub H)P(+15.99)GTVGPAGVR	577.3196	2	1071	H → G	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(sub G)DAGPP(+15.99)GPAGPTGPP(+15.99)GPIGNVAGAP(+15.99)GP(+15.99)K	1205.582	2	833	G → N	COL1A1	XP_026368914.1	<i>Ursus arctos horribilis</i>
	N(sub G)ESGNKGEP(+15.99)GSVGPQGGPP(+15.99)GPSGEEGK	1241.0632	2	356	G → N	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	Q(sub A)GEK(+15.99)GHAGLAGAR	635.3124	2	508	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(sub G)ETGLRGDVGNP(+15.99)GR	486.5744	3	659	G → N	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
Sc-6	GEND(sub G)PVGPTGPVGSAGPSGPN(+.98)GPP(+15.99)GPAGSR	886.7492	3	758	G → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	TN(sub G)ETGAHGPP(+15.99)GFAGEK	793.8668	2	833	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(sub G)PP(+15.99)GAVGAP(+15.99)GVN(+.98)GAP(+15.99)GEAGR	897.9244	2	908	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	D(sub H)GN(+.98)RGEP(+15.99)GPAGAVGPVAVGPR	502.0032	4	977	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	AGEE(sub K)GHAGLAGAR	399.1975	3	511	K → E	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	AGEKGD(sub H)AGLAGAR	586.8041	2	513	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
Sc-7	GM(sub E)NGPVGPTGPVGSAGPSGPN(+.98)GPN(sub P)GPAGSR	1302.1116	2	722 750	M → E P → N	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	TGEG(sub T)GAHGPPGY(sub F)AGEK	742.8411	2	835 843	G → T F → Y	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	GPAGDP(+15.99)GKN(+.98)GDKGD(sub H)AGLAGPR	962.4590	2	512	H → D	COL1A2	XP_010585560.1	<i>loxodonta africana</i>
	GEVGPAGPN(+.98)GFAGPAGAAGQ(+.98)AGAE(sub K)GER	785.3611	3	738	K → E	COL1A2	XP_010585560.1	<i>loxodonta africana</i>
Sc-8	TN(sub G)ETGASGPP(+15.99)GFAGEK	768.3513	2	832	G → N	COL1A2	XP_010585560.1	<i>loxodonta africana</i>
	TGETGAD(sub S)GPP(+15.99)GFAGEK	753.8379	2	837	S → D	COL1A2	XP_010585560.1	<i>loxodonta africana</i>
	GPAGE(sub P)SGPVGK	478.2436	2	1058	G → N	COL1A2	XP_010585560.1	<i>loxodonta africana</i>
	GSAGPD(sub P)GATGFP(+15.99)GAAGR	731.3383	2	870	P → D	COL1A1	NP_001029211.1	<i>Bos primigenius taurus</i>
Sc-9	GGP(+15.99)GSE(sub R)GFP(+15.99)GADGVAGPK	794.8610	2	492	R → E	COL1A1	NP_001029211.1	<i>Bos primigenius taurus</i>

*Extinct species identification from Upper Pleistocene bone fragments
not identifiable from their osteomorphological studies by proteomics analysis*

	GFP(+15.99)GADGVAGPE(sub K)GPAGER	828.8846	2	537	K → E	COL1A1	NP_001029211.1	<i>Bos primigenius taurus</i>
	N(sub G)PP(+15.99)GESGAAGPTGPIGSR	819.3964	2	590	G → N	COL1A2	NP_776945.1	<i>Bos primigenius taurus</i>
	D(sub H)GN(+.98)RGEP(+15.99)GPAGAVGPAGAVGPR	988.4811	2	973	H → D	COL1A2	NP_776945.1	<i>Bos primigenius taurus</i>
	N(sub G)DAGPP(+15.99)GPAGPTGPP(+15.99)GPIGNVVGAP(+15.99)GPK	1197.592	2	833	G → N	COL1A1	XP_026368914.1	<i>Ursus arctos horribilis</i>
	RGPN(+.98)GEAGSAGPSGPP(+15.99)GLE(sub R)GSP(+15.99)GSR	761.0187	3	400	R → E	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	AGEE(sub K)GHAGLAGAR	399.1973	3	511	K → E	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	N(sub G)PP(+15.99)GAVGAP(+15.99)GVNGAP(+15.99)GEAGR	896.9321	2	908	G → N	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
Sc-10	D(sub H)GN(+.98)RGEP(+15.99)GPAGAVGPVAVGPR	501.7508	4	976	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	D(sub G)NRGEP(+15.99)GPAGAVGPVAVGPR	649.3326	3	977	G → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	GPSGPQ(+.98)GVRGDE(sub K)GEP(+15.99)GDKGPR	689.6580	3	1009	K → E	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>
	IGD(sub H)P(+15.99)GTVGPAGVR	577.3197	2	1071	H → D	COL1A2	XP_026361636.1	<i>Ursus arctos horribilis</i>

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

The samples Sc-1 to Sc-3 show twelve mutations on the sequence of collagens referencing to the *Panthera pardus* species. On the COL1A1 sequence, the substitution of arginine (R) by glutamic acid (E) is observed at position 793, the glycine (G) by asparagine (N) at position 779, the alanine (A) by glutamine (Q) at position 781 and the glycine (G) by asparagine (N) at position 1067. On the COL1A2 sequence, eight substitutions are observed at the positions 513, 833, 835, 908, 976, 977, 998 and 1071 (Table 5). On sample Sc-4, four mutations were detected; 3 on the COL1A1 sequence on position 793, 838 and 1063 and the last on the position 513 of the COL1A2 sequence (Table 5). The mutation on the COL1A1 sequence on position 793 and on the COL1A2 sequence on position 513 are also detected for samples Sc-1 to Sc-3 whereas the sample Sc-4 carries no other mutation. In brief samples Sc-1 to Sc-3 carries 10 identical amino acid variations in their protein sequences showing their belonging to the same species. These observations are in agreement with the osteomorphological analyses concluding to the identification of the *Panthera spelaea* species for the 3 samples, Sc-1 to Sc-3. The sample Sc-4 carries only 4 mutations in the collagen 1 alpha 1 and alpha 2 sequences compared to the modern *Panthera pardus* sequences. This observation is also in agreement with the identification made by the paleontologists.

The Sc-5 sample has been identified as a bone from the *Ursus spelaeus* species by paleontologists. The closest ancestor sequence in the databases is *Ursus arctos horribilis*. Proteomic analysis shows one amino acid substitution on the sequence of COL1A1 and eleven amino acid substitutions on the sequence of COL1A2 (Table 5). The Sc-6 and Sc-7 samples were identified as *Ursus thibetanus* species bones by paleontologists. Its closest ancestor is again *Ursus arctos horribilis*. One amino acid substitution on the sequence of COL1A1 and seven amino acid substitutions on the sequence of COL1A2 in Sc-6 sample. Four mutations in Sc-7 were observed on the sequence of COL1A2. There are no common mutations for samples Sc-6 and Sc-7. Only the mutations on the position 511 and 513 are common to the samples Sc-7 and Sc-5.

For the sample Sc-8, the closest ancestor is *Loxodonta Africana*. No substitution is observed on COL1A1 but 5 substitutions are observed on COL1A2 (Table 5).

The Sc-9 retoucher is identified as belonging to the *Bovidae* family. Considering the Scladina excavation site fauna, the closest animals are *Bos primigenius primigenius* (aurochs) and *Bison priscus* (bison). Collagen sequence of *Bos primigenius taurus* was identified with a greater number of peptides and a slightly higher sequence coverage. For COL1A1, *Bos primigenius taurus*: 408 peptides, 69% of coverage; *Bison priscus*: 149 peptides; 66% of coverage. For COL1A2, *Bos taurus*: 384 peptides, 71% of coverage; *Bison bison*: 243 peptides, 71% of coverage.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Three amino acid substitutions were also identified on the COL1A1 sequence and 2 substitutions on COL1A2 referring to *Bos Taurus* species. These results tend to show that the Sc-9 sample is a *Bos primigenius taurus* (aurochs) in agreement with Scladina Cave fauna.

Proteomics data indicates that the Sc-10 retoucher has been made from a bone of the *Ursidae* family. The closest species in protein database, referring to the Scladina Cave fauna is *Ursus arctos horribilis* as discussed above. Proteomics detected 1 mutations on COL1A1 sequence 7 mutations on COL1A2 sequence. Seven of these mutations are common with the Sc-5 sample (833, 511, 908, 976, 977, 1009 and 1071 positions). These results show a resemblance between the samples Sc-10 and Sc-5 potentially concluding that the Sc-10 sample may belong to the *Ursus spelaeus* species. The comparison between Sc-10 and Sc-6 and 7, show two mutation is in common with Sc-6 identified as at positions 833, 977 and one mutation is in common with Sc-7 identified as at positions 511. This results concluding that the samples Sc-6 and Sc-7 are not belonging to the same species than the sample Sc-10.

4. Discussion

4.1 Protein identification

Bone preservation and the presence of modern contaminants are two major difficulties of bone palaeoproteomics. In our study, the keratins were identified in minority in comparison to endogenous bone proteins like collagens. Human honerin, dermicin which are very often described as contaminants in bones proteomic analysis have been detected (Sawafuji, et al., 2017, Wadsworth and Buckley, 2014).

The two chains of the type I collagen, that represents 90% of the bone organic matter, were identified with up to 60% coverage for both chains. This result shows an exceptional bone preservation and a highly sensitive proteomic analysis based on an eFASP protocol (Erde, et al., 2014) without protein extraction after demineralization and 5 mg starting material. It can be pointed out that starting material ranging from 30 mg to several hundred mg are commonly referenced in palaeoproteomics studies (Buckley, et al., 2017a, Buckley, et al., 2017b, Buckley, et al., 2019, Cappellini, et al., 2011, Welker, et al., 2017)).

In this study, NCPs are not detected despite described in several ancient bones studies (Cappellini, et al., 2011, Sawafuji, et al., 2017, Wadsworth and Buckley, 2014) have been identified in our study. It is important to note that other collagens have been identified as the COL2A1 which is a fibrillar collagen found in cartilage, the COL3A1 which is found as a major structural component in hollow organs and the COL5A2 which is a minor component of connective tissue. Welker and

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

colleagues showed the presence of others collagens as COL10A1 and COL27A1 in bones specimens from the Grotte du Renne (Welker, et al., 2016).

The LC-MS/MS of paleontological bone optimized eFASP digestion afford a high number of identified peptides, an important percentage of coverage and the identification of PTMs as very few spurious modifications are induced. The phylogenetic information on the species from the Scladina Cave afforded by paleontologists allow to identify the closest species in modern protein databases. By combining proteomic analysis, phylogenetic information and modern species protein data bases we identify the species of bones from the Scladina Cave which are in agreement with the paleontologists' data. This methodology was applied to two retouchers Sc-9 and Sc-10 the species of which cannot be identified by their anatomy. Our study works on 5 mg of bone powder which is less than previously used quantities. This quantity is sufficient to identified proteins, genus, PTMs from Pleistocene bones, which indicates that our method is extremely sensitive.

4.2 PTMs

The preservation of bone proteins may be deduced from the percentage in weight of extracted collagen and the rate of deamidation which has been reported as a marker of degradation in ancient bone (Van Doorn, et al., 2012). Deamidation ranges from 0 to 20% for bones from 0 to 2000 years while older fossils show an extreme variance in their total levels of deamidation (Schroeter and Cleland, 2016). Deamidations of around 50% of glutamines and 60% of asparagine were found for 40 and 50 ka bones from the Kleine Feldhofer Grotte (Germany) (Lanigan, et al., 2020). Considering middle Pleistocene bone specimens of the rhinoceros genus *Stephanorhinus* (337–300 ka), glutamine and asparagine deamidations reach nearly 100% (Welker, et al., 2017). In our study, the rate of deamidation ranges between 71 – 95% on the COL1A1 and 71 – 92% on the COL1A2 specific peptides identified in all Sc1-10 samples which is in agreement to the deamidation rates described in the literature (Welker, et al., 2017). The percentage of deamidation we observed in the demineralization residual fraction of collagen I alpha 1 is very close to the one observed in the bone power fraction.

Regarding proline oxidation, the collagen identified shows many peptides with hydroxylated prolines. The percentage of frequency is close to 99%, on the 60% of the collagen sequence which has been identified. In modern samples, the collagen prolines are oxidized at 20% (Zaitseva, et al., 2015). This modification is involved in stabilizing triple helix of collagen (Kotch, et al., 2008). The presence of peptides with hydroxyprolines can be explained by the age of the bones and the state of conservation.

It is also interesting to notice that hydroxylysines were identified in the collagen sequences of Scladina Cave bones. Hydroxylysines are involved in crosslinks between the triple helix of collagen (Knott and Bailey, 1998). Hydroxylysines can be further modified by the sequential steps of *O*-linked

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

glycosylation producing G-Hyl (galactosylhydroxylysine) and GG-Hyl (glucosylgalactosylhydroxylysine) (Scott, et al., 2012) and play multiple roles in normal mammalian physiology and pathology. Ryan C. Hill et al., have shown the presence of hydroxylysine glucosylgalactosylation in the extinct *Bison latifrons* (Hill, et al., 2015). Dihydroxyproline were identified in Scaldina cave samples. This modification has been reported in siliceous cell walls of diatoms (Ehrlich, et al., 2010, Nakajima and Volcani, 1969, Surmik, et al., 2016). These authors suggested that hydroxylated amino acids could play a role in silicification of diatom cell walls. The silicification is a fossilization process whereby the organism is penetrated by silica that form on the cells and cell structures. Bonjean and colleague showed the presence of silicified materials in the area of the Scaldina Cave, which could explain the presence of dihydroxylation of prolines on the collagen sequences (Bonjean, et al., 2015).

4.3 Molecular sequence variation

Ancient proteins are commonly identified in references to the closest (most often modern) species present in the database as shown for Pleistocene samples of *Raphus cucullatus* from modern birds (Horn, et al., 2019), *Stephanorhinus* from *Rhinoceros* (Cappellini, et al., 2019, Welker, et al., 2017) or kangaroos from modern Australian vertebrate species (echidna, wombat, red kangaroo...) (Buckley, et al., 2017a).

This study shows a successful identification of bone taxa from Scaldina Cave using proteomics and paleontological data analysis. Among the taxa identified by proteomics are the Felidae family, Elephantidae family, the Bovinae subfamily, the *Ursus* genus. Amino acid substitutions on the collagens were identified providing new information on extinct species sequences and helping in taxonomy-based clustering.

The samples Sc-1 to Sc-3 were identified as belonging to the Felidae family. No species information was extracted from our dataset. Osteomorphological studies identified the *Panthera spelaea* species however, no collagens are referenced in the database for this species (only mitochondrial sequences) nor for the closest species *Panthera leo* (Barnett, et al., 2016). The second closest species is *Panthera pardus* but again, no discriminant peptide was identified for this species nor the other referenced species (*Acinonyx jubatus*, *Felis silvestris catus*, *Lynx pardinus*, *Panthera pardus*, *Panthera tigris altaica*, *Puma concolor*). Same amino acid substitutions of the COL1A1 and COL1A2 sequenced were found for the 3 samples (by referenced to the *Panthera pardus* modern sequence) indicating their belonging to the same species. A fourth sample, Sc-4, identified as belonging to the Felidae family showed different collagen sequence mutations compared to Sc-1 to Sc-3 samples indicating its belonging to another species. The Sc-4 bone has been identified as *Panthera pardus* by the paleontologists. These observations are in agreement with the phylogenetic studies

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

carried out on mitochondrial DNA of *Panthera spelaea* (Barnett, et al., 2016, Joachim, et al., 2004, Ma and Wang, 2015).

The samples Sc-5, Sc-6, Sc-7 were identified as belonging to the Ursidae family by proteomics. Osteomorphological studies identified the Sc-5 sample as belonging to the *Ursus spelaeus* and the Sc-6 and Sc-7 samples to *Ursus thibetanus* species. The closest ancestor sequence in the databases is *Ursus arctos horribilis* for both species. Amino acids variations identified for the 3 samples could not help in taxonomy-based clustering. For the two bones Sc-6 and Sc-7, there is no mutations in common but the paleontologists have identified by osteomorphological studies as being of the same species. This can show variability between the collagen protein sequences of the same species. It would be interesting to have the genome sequencing of the species *Ursus thibetanus* species to confirm the Osteomorphological studies.

The sample Sc-8 showed specificity to the *Mammuth americanum* species using 2 discriminant peptides on COL1A2 sequence and other unique peptides correspond to the Proboscidea order. The identification of the Proboscidea order taxa classification is in full agreement with the osteomorphological identification realized by the palaeontologists. The specific peptides, which showed specificity to the *Mammuth americanum* species, are in contradiction with osteomorphological and faunal studies. We show in this study of importance of correlating the proteomics data with the palaeontologists studies and to work on the protein sequences of the closest species.

The retouchers (Sc-9 and Sc-10) could not be identified by osteomorphologic studies. Proteomics identified the Sc-9 retoucher as belonging to the *Bovidae* family with a higher responses to *Bos primigenius primigenius* (aurochs) (number of peptides, slightly higher sequence coverage) in agreement with Scladina Cave fauna. The Sc-10 retoucher was identified to belong to the genus *Ursus*. Whereas Sc-10 shows different mutations than those observed for the Sc-6 and Sc-7 samples, similarities in amino acid mutations to the Sc-5 sample tend to conclude a to the Sc-10 belonging to the *Ursus spelaeus* species. Abrams *et al.* highlighted the presence of bear bone retouchers used by the local Neanderthal community present on the Scladina site (Abrams, et al., 2014) which were used tools for lithic activities. This taxa identification helps to confirm the exploitation of bear carcasses (brown and cave bears), recognised by the presence of butchery marks which remains a rare event during the Middle Palaeolithic. The use of bear bones by Neandertals is not very current and few sites shows remains of brown or cave bears concerned as retouchers, as in Northern France at Biache-saint-Vaast (Moigne, et al., 2016). So, this protocol of analysis, with a very low quantity of bone used, appear to be of great interest for Palaeolithic prehistorians.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

Overall, the taxa identified by proteomics for the 10 studied samples correlate with the osteomorphological, palaeoenvironmental and palaeodietary data (Bocherens, et al., 1999, Bocherens, et al., 1997).

5. Conclusion

Our study has demonstrated the possibility to identify taxa of 130,000 years old bones using an optimized proteomics method starting from 5 mg of samples. A large number of peptides was identified along with high average sequence coverages up to of 60% for collagen 1 alpha 1. The deamidation frequency of about 80% and oxidation frequency of about 99% were measured. Despite the restricted databases for several species, several taxa were identified including the Felidae family, Elephantidae, Bovinae subfamily, the *Ursus* genus. Our study has shown not referenced amino acids substitutions specific to the studied species. Proteomics allowed to identify the retouchers' originating families, bovidae and ursidae at Scladina Cave at the beginning of the Upper Pleistocene by Neandertals. Considering the amino acid variations of the collagen sequences and paleontological studies, one retoucher was identified as potentially belonging to the *Ursus spelaeus* species.

6. Declaration of Competing Interest

The authors declare there is no conflict of interest.

7. Acknowledgments

The authors acknowledge the IBISA network for financial support of the USR 3290 (MSAP) proteomics facility. The mass spectrometers were funded by the University of Lille, the CNRS, the Région Hauts-de-France and the European Regional Development Fund. The authors thank the CNRS Mission pour l'Interdisciplinarité for the funding of the project ProtéoMamQuat (2017-2018). The authors warmly thank the Scladina cave excavation team for granting them the bone samples of this study and for sharing their palaeontological data.

8. Credit roles

Fabrice Bray, Christian Rolando, Caroline Tokarski, Patrick Auguste: Conceptualization, Methodology. Dominique Bonjean, Grégory Abrams, Kévin Di Modica, Patrick Auguste: Resources. Fabrice Bray, Stéphanie Flament: Investigation. Fabrice Bray, Christian Rolando, Caroline Tokarski, Patrick Auguste: Writing – Original Draft. Fabrice Bray, Dominique Bonjean, Grégory Abrams, Kévin Di Modica, Christian Rolando, Caroline Tokarski, Patrick Auguste: Writing – Review & Editing.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

9. Appendix A. Supplementary data

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD021171.

The supplementary data 1 corresponds to pictures of ten bones from the archaeological excavation of Scladina Cave. The supplementary data 2 are excel files which containing the identification of proteins with Mascot engine against the Mammalian database. The supplementary data 2 contains one excel file for protein identified in demineralizing fraction and one other file for protein identified in powder fraction. The supplementary data 3 corresponds to the fragmentation spectra of unique peptides identified and presented in table 3. The supplementary data 4 corresponds to excel files containing the identified peptides for ten bones and both fractions from the archaeological excavation of Scladina Cave. The supplementary data 5 corresponds to the fragmentation spectra of peptides with oxidation on methionine, proline, asparagine and glutamine deamidation. Supplementary data 5 contains the figures of deamidation and oxidation frequencies for peptides form COL1A1 and COL1A2 for both fractions. Supplementary data 6 corresponds to the fragmentation spectra of peptides with amino acids variations identified in the studied archaeological samples.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

10. References

- Abrams, G., Bello, S.M., Di Modica, K., Pirson, S., Bonjean, D., (2014). When Neanderthals used cave bear (*Ursus spelaeus*) remains: Bone retouchers from unit 5 of Scladina Cave (Belgium), *Quat. Int.* 326-327, 274-287, <https://doi.org/10.1016/j.quaint.2013.10.022>.
- Abrams, G., Bonjean, D., Di Modica, K., Pirson, S., Otte, M., Patou-Mathis, M., (2010). Les os brûlés de l'ensemble sédimentaire 1A de Scladina (Andenne, Belgique): apports naturels ou restes de foyer (s) néandertalien (s), *Notae Praehistoricae*, 5-13.
- Barnett, R., Mendoza, M.L.Z., Soares, A.E.R., Ho, S.Y.W., Zazula, G., Yamaguchi, N., Shapiro, B., Kirillova, I.V., Larson, G., Gilbert, M.T.P., (2016). Mitogenomics of the Extinct Cave Lion, *Panthera spelaea* (Goldfuss, 1810), Resolve its Position within the Panthera Cats, *Open Quat.* 2, <https://doi.org/10.5334/oq.24>.
- Birch, S.E.P., Scheu, A., Buckley, M., Çakırlar, C., (2019). Combined osteomorphological, isotopic, aDNA, and ZooMS analyses of sheep and goat remains from Neolithic Ulucak, Turkey, *Archaeol. Anthropol. Sci.* 11, 1669-1681, <https://doi.org/10.1007/s12520-018-0624-8>.
- Bocherens, H., Billiou, D., Mariotti, A., Patou-Mathis, M., Otte, M., Bonjean, D., Toussaint, M., (1999). Palaeoenvironmental and palaeodietary implications of isotopic biogeochemistry of last interglacial Neanderthal and mammal bones in Scladina Cave (Belgium), *J. Archaeol. Sci.* 26, 599-607, <https://doi.org/10.1006/jasc.1998.0377>.
- Bocherens, H., Billiou, D., Patou-Mathis, M., Bonjean, D., Otte, M., Mariotti, A., (1997). Paleobiological implications of the isotopic signatures (¹³C, ¹⁵N) of fossil mammal collagen in Scladina Cave (Sclayn, Belgium), *Quat. Res.* 48, 370-380, <https://doi.org/10.1006/qres.1997.1927>.
- Bonjean, D., Abrams, G., Di Modica, K., Otte, M., (2009). La microstratigraphie, une clé de lecture des remaniements sédimentaires successifs. Le cas de l'industrie moustérienne 1A de Scladina, *Notae Praehistoricae*, 139-147.
- Bonjean, D., Vanbrabant, Y., Abrams, G., Pirson, S., Burlet, C., Di Modica, K., Otte, M., Vander Auwera, J., Golitko, M., McMillan, R., Goemaere, E., (2015). A new Cambrian black pigment used during the late Middle Palaeolithic discovered at Scladina Cave (Andenne, Belgium), *J. Archaeol. Sci.* 55, 253-265, <https://doi.org/10.1016/j.jas.2014.11.040>.
- Brandt, L.Ø., Haase, K., Collins, M.J., (2018). Species identification using ZooMS, with reference to the exploitation of animal resources in the medieval town of Odense, *Dan. J. Archaeol.* 7, 139-153, <https://doi.org/10.1080/21662282.2018.1468154>.
- Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., Derevianko, A., Buckley, M., (2016). Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis, *Sci. Rep.* 6, 23559, <https://doi.org/10.1038/srep23559>.
- Buckley, M., (2015). Ancient collagen reveals evolutionary history of the endemic South American 'ungulates', *Proc. R. Soc. B* 282, 20142671, <https://doi.org/10.1098/rspb.2014.2671>.
- Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C.W., (2009). Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* 23, 3843-3854, <https://doi.org/10.1002/rcm.4316>.
- Buckley, M., Cosgrove, R., Garvey, J., Prideaux, G.J., (2017a). Identifying remains of extinct kangaroos in Late Pleistocene deposits using collagen fingerprinting, *J. Quat. Sci.* 32, 653-660, <https://doi.org/10.1002/jqs.2964>.
- Buckley, M., Harvey, V.L., Chamberlain, A.T., (2017b). Species identification and decay assessment of Late Pleistocene fragmentary vertebrate remains from Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting, *Boreas* 46, 402-411, <https://doi.org/10.1111/bor.12225>.
- Buckley, M., Lawless, C., Rybczynski, N., (2019). Collagen sequence analysis of fossil camels, *Camelops* and c.f. *Paracamelus*, from the Arctic and sub-Arctic of Plio-Pleistocene North America, *J. Proteomics* 194, 218-225, <https://doi.org/10.1016/j.jprot.2018.11.014>.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

- Cappellini, E., Gentry, A., Palkopoulou, E., Ishida, Y., Cram, D., Roos, A.-M., Watson, M., Johansson, U.S., Fernholm, B., Agnelli, P., (2014). Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae), *Zool. J. Linn. Soc.* 170, 222-232, <https://doi.org/10.1111/zoj.12084>.
- Cappellini, E., Jensen, L.J., Szklarczyk, D., Ginolhac, A., da Fonseca, R.A., Stafford Jr, T.W., Holen, S.R., Collins, M.J., Orlando, L., Willerslev, E., (2011). Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins, *J. Proteome Res.* 11, 917-926, <https://doi.org/10.1021/pr200721u>.
- Cappellini, E., Prohaska, A., Racimo, F., Welker, F., Pedersen, M.W., Allentoft, M.E., de Barros Damgaard, P., Gutenbrunner, P., Dunne, J., Hammann, S., (2018). Ancient biomolecules and evolutionary inference, *Annu. Rev. Biochem.* 87, 1029-1060, <https://doi.org/10.1146/annurev-biochem-062917-012002>.
- Cappellini, E., Welker, F., Pandolfi, L., Ramos-Madrigal, J., Samodova, D., R  ther, P.L., Fotakis, A.K., Lyon, D., Moreno-Mayar, J.V., Bukhsianidze, M., Jersie-Christensen, R.R., Mackie, M., Ginolhac, A., Ferring, R., Tappen, M., Palkopoulou, E., Dickinso, M.R., Stafford Jr., T.W., Chan, Y.L., G  therstr  m, A., Nathan, S.K.S.S., Heintzman, P.D., Kapp, J.D., Kirillova, I.V., Moodley, Y., Agusti, J., Kahlke, R.-D., Kiladze, G., Mart  nez-Navarro, B., Liu, S., Velasco, M.S., Sinding, M.-H.S., Kelstrup, C.D., Allentoft, M.E., Orlando, L., Penkman, K., Shapiro, B., Rook, L., Dal  n, L., Gilbert, M.T.P., Olsen, J.V., Lordkipanidze, D., Willerslev, E., (2019). Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny, *Nature* 574, 103-107, <https://doi.org/10.1038/s41586-019-1555-y>.
- Cleland, T.P., (2018a). Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics, *J. Proteome Res.* 17, 3976-3983, <https://doi.org/10.1021/acs.jproteome.8b00637>.
- Cleland, T.P., (2018b). Solid Digestion of Demineralized Bone as a Method To Access Potentially Insoluble Proteins and Post-Translational Modifications, *J. Proteome Res.* 17, 536-542, <https://doi.org/10.1021/acs.jproteome.7b00670>.
- Cleland, T.P., Schroeter, E.R., (2018). A comparison of common mass spectrometry approaches for paleoproteomics, *J. Proteome Res.* 17, 936-945, <https://doi.org/10.1021/acs.jproteome.7b00703>.
- Cleland, T.P., Schroeter, E.R., Feranec, R.S., Vashishth, D., (2016). Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics, *Proc. Biol. Sci.* 283, 20160593, <https://doi.org/10.1098/rspb.2016.0593>.
- Dallongeville, S., Garnier, N., Rolando, C., Tokarski, C., (2016). Proteins in art, archaeology, and paleontology: from detection to identification, *Chem. Rev.* 116, 2-79, <https://doi.org/10.1021/acs.chemrev.5b00037>.
- Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B.M., (2016). Protein sequences bound to mineral surfaces persist into deep time, *elife* 5, e17092, <http://doi.org/10.7554/eLife.17092>.
- Desmond, A., Barton, N., Bouzouggar, A., Douka, K., Fernandez, P., Humphrey, L., Morales, J., Turner, E., Buckley, M., (2018). ZooMS identification of bone tools from the North African Later Stone Age, *J. Archaeo.Sci.* 98, 149-157, <https://doi.org/10.1016/j.jas.2018.08.012>.
- Ehrlich, H., Deutzmann, R., Brunner, E., Cappellini, E., Koon, H., Solazzo, C., Yang, Y., Ashford, D., Thomas-Oates, J., Lubeck, M., (2010). Mineralization of the metre-long biosilica structures of glass sponges is templated on hydroxylated collagen, *Nat. chem.* 2, 1084-1088, <https://doi.org/10.1038/nchem.899>
- Erde, J., Ogorzalek Loo, R.R., Loo, J.A., (2014). Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments, *J. Proteome Res.* 13, 1885-1895, <https://doi.org/10.1021/pr4010019>.
- Harvey, V.L., LeFebvre, M.J., deFrance, S.D., Toftgaard, C., Drosou, K., Kitchener, A.C., Buckley, M., (2019). Preserved collagen reveals species identity in archaeological marine turtle bones from Caribbean and Florida sites, *Royal Soc. Open Sci.* 6, 191137, <https://doi.org/10.1098/rsos.191137>.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

- Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C., Collins, M.J., (2018). A guide to ancient protein studies, *Nat. Ecol. Evol.* 2, 791-799, <https://doi.org/10.1038/s41559-018-0510-x>.
- Hill, R.C., Wither, M.J., Nemkov, T., Barrett, A., D'Alessandro, A., Dzieciatkowska, M., Hansen, K.C., (2015). Preserved Proteins from Extinct Bison latifrons Identified by Tandem Mass Spectrometry; Hydroxylysine Glycosides are a Common Feature of Ancient Collagen, *Mol. Cell. Proteomics* 14, 1946-1958, <https://doi.org/10.1074/mcp.M114.047787>.
- Hofman, C.A., Rick, T.C., Erlandson, J.M., Reeder-Myers, L., Welch, A.J., Buckley, M., (2018). Collagen Fingerprinting and the Earliest Marine Mammal Hunting in North America, *Sci. Rep.* 8, 10014, <https://doi.org/10.1038/s41598-018-28224-0>.
- Horn, I.R., Kenens, Y., Palmblad, N.M., Van der Plas-Duivesteyn, S.J., Langeveld, B.W., Meijer, H.J.M., Dalebout, H., Marissen, R.J., Fischer, A., Florens, V.F.B., Niemann, J., Rijdsdijk, K.F., Schulp, A.S., Laros, J.F.J., Gravendeel, B., (2019). Palaeoproteomics of bird bones for taxonomic classification, *Zool. J. Linn. Soc.* 186, 650-665, <https://doi.org/10.1093/zoolinnean/zlz012>.
- Joachim, B., Wilfried, R., Odile, L., Helmut, H., Torsten, E., Anders, G., Jennifer, H., Matthew, J.C., Timothy, W., Kurt, W.A., (2004). Molecular phylogeny of the extinct cave lion *Panthera leo spelaea*, *Molecular phylogenetics and evolution* 30, 841-849, 10.1016/j.ympev.2003.07.020.
- Knott, L., Bailey, A.J., (1998). Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance, *Bone* 22, 181-187, [https://doi.org/10.1016/S8756-3282\(97\)00279-2](https://doi.org/10.1016/S8756-3282(97)00279-2).
- Kostyukevich, Y., Bugrova, A., Chagovets, V., Brzhozovskiy, A., Indeykina, M., Vanyushkina, A., Zherebker, A., Mitina, A., Kononikhin, A., Popov, I., Khaitovich, P., Nikolaev, E., (2018). Proteomic and lipidomic analysis of mammoth bone by high-resolution tandem mass spectrometry coupled with liquid chromatography, *Eur. J. Mass Spectrom.* 24, 411-419, <https://doi.org/10.1177/1469066718813728>.
- Kotch, F.W., Guzei, I.A., Raines, R.T., (2008). Stabilization of the collagen triple helix by O-methylation of hydroxyproline residues, *J. Am. Chem. Soc.* 130, 2952-2953, <https://doi.org/10.1021/ja800225k>.
- Lanigan, L.T., Mackie, M., Feine, S., Hublin, J.-J., Schmitz, R.W., Wilcke, A., Collins, M.J., Cappellini, E., Olsen, J.V., Taurozzi, A.J., Welker, F., (2020). Multi-protease analysis of Pleistocene bone proteomes, *J. Proteomics* 228, 103889, <https://doi.org/10.1016/j.jprot.2020.103889>.
- Ma, Y.-p., Wang, S., (2015). Mitochondrial genome of the African lion *Panthera leo leo*, *Mitochondrial DNA* 26, 951-952, <https://doi.org/10.3109/19401736.2013.865171>.
- Martisius, N.L., Welker, F., Dogandžić, T., Grote, M.N., Rendu, W., Sinet-Mathiot, V., Wilcke, A., McPherron, S.J., Soressi, M., Steele, T.E., (2020). non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals, *Sci. Rep.* 10, 1-12, <https://doi.org/10.1038/s41598-020-64358-w>.
- McGrath, K., Rowsell, K., Gates St-Pierre, C., Tedder, A., Foody, G., Roberts, C., Speller, C., Collins, M.J., (2019). Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points, *Sci. Rep.* 9, 11027, <https://doi.org/10.1038/s41598-019-47299-x>.
- Moigne, A.-M., Valensi, P., Auguste, P., García-Solano, J., Tuffreau, A., Lamotte, A., Barroso, C., Moncel, M.-H., (2016). Bone retouchers from Lower Palaeolithic sites: Terra Amata, Orgnac 3, Cagny-l'Épinette and Cueva del Angel, *Quat. Int.* 409, 195-212, <https://doi.org/10.1016/j.quaint.2015.06.059>.
- Nakajima, T., Volcani, B., Elzari, (1969). 3,4-dihydroxyproline: a new amino acid in diatom cell walls, *Science* 164, 1400-1401, <https://doi.org/10.1126/science.164.3886.1400>.
- Nielsen-Marsh, C.M., Ostrom, P.H., Gandhi, H., Shapiro, B., Cooper, A., Hauschka, P.V., Collins, M.J., (2002). Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka, *Geol.* 30, 1099-1102, [https://doi.org/10.1130/0091-7613\(2002\)030<1099:SPOOPA>2.0.CO;2](https://doi.org/10.1130/0091-7613(2002)030<1099:SPOOPA>2.0.CO;2).
- Nielsen-Marsh, C.M., Richards, M.P., Hauschka, P.V., Thomas-Oates, J.E., Trinkaus, E., Pettitt, P.B., Karavanić, I., Poinar, H., Collins, M.J., (2005). Osteocalcin protein sequences of Neanderthals and modern primates, *Proc. Natl. Acad. Sci.* 102, 4409-4413, <https://doi.org/10.1073/pnas.0500450102>

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

- Ostrom, P.H., Gandhi, H., Strahler, J.R., Walker, A.K., Andrews, P.C., Leykam, J., Stafford, T.W., Kelly, R.L., Walker, D.N., Buckley, M., (2006). Unraveling the sequence and structure of the protein osteocalcin from a 42 ka fossil horse, *Geochim. Cosmochim. Acta* 70, 2034-2044, <https://doi.org/10.1016/j.gca.2006.01.004>.
- Ostrom, P.H., Schall, M., Gandhi, H., Shen, T.-L., Hauschka, P.V., Strahler, J.R., Gage, D.A., (2000). New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry, *Geochim. Cosmochim. Acta* 64, 1043-1050, [https://doi.org/10.1016/S0016-7037\(99\)00381-6](https://doi.org/10.1016/S0016-7037(99)00381-6).
- Otte, M., Léotard, J.-M., Schneider, A.-M., Gautier, A., (1983). Fouilles aux grottes de Sclayn (Namur), *Helinium Wetteren* 23, 112-142.
- Otte, M., Patou-Mathis, M., Bonjean, D., (2000). Les espèces chassées et consommées par l'homme en couche 5, in: Tuffreau, A. (Ed.), Bulletin de la Société préhistorique française, *Université de Liège Liège*, pp. 669-670.
- Park, S.D., Magee, D.A., McGettigan, P.A., Teasdale, M.D., Edwards, C.J., Lohan, A.J., Murphy, A., Braud, M., Donoghue, M.T., Liu, Y., Chamberlain, A.T., Rue-Albrecht, K., Schroeder, S., Spillane, C., Tai, S., Bradley, D.G., Sonstegard, T.S., Loftus, B.J., MacHugh, D.E., (2015). Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle, *Genome Biol.* 16, 234, <https://doi.org/10.1186/s13059-015-0790-2>.
- Pirson, S., Court-Picon, M., Haesaerts, P., Bonjean, D., Damblon, F., (2008). New data on geology, anthracology and palynology from the scladina cave Pleistocene sequence preliminary results, *Memoirs of the Geological Survey of Belgium* 55, 71-93.
- Pirson, S., Damblon, F., Balescu, S., Bonjean, D., Haesaerts, P., (2014). The palaeoenvironmental context and chronostratigraphic framework of the Scladina Cave sedimentary sequence (units 5 to 3-SUP), *Palaeoanthropology and Context ed., Etudes et Recherches archéologiques de l'Université de Liège*, 134, 69-92.
- Presslee, S., Slater, G.J., Pujos, F., Forasiepi, A.M., Fischer, R., Molloy, K., Mackie, M., Olsen, J.V., Kramarz, A., Taglioretti, M., (2019). Palaeoproteomics resolves sloth relationships, *Nat. Ecol. Evol.* 3, 1121-1130, <https://doi.org/10.1038/s41559-019-0909-z>.
- Ramsøe, A., van Heekeren, V., Ponce, P., Fischer, R., Barnes, I., Speller, C., Collins, M.J., (2020). DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes, *J. Archaeo.Sci.* 115, 105080, <https://doi.org/10.1016/j.jas.2020.105080>.
- Rohland, N., Malaspinas, A.-S., Pollack, J.L., Slatkin, M., Matheus, P., Hofreiter, M., (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup, *PLOS Biol.* 5, 207, <https://doi.org/10.1371/journal.pbio.0050207>.
- 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, *Royal Soc. Open Sci.* 4, 161004, <https://doi.org/10.1098/rsos.161004>.
- Schroeter, E.R., Blackburn, K., Goshe, M.B., Schweitzer, M.H., (2019). Proteomic method to extract, concentrate, digest and enrich peptides from fossils with coloured (humic) substances for mass spectrometry analyses, *Royal Soc. Open Sci.* 6, 181433, <https://doi.org/10.1098/rsos.181433>.
- 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., Cleland, T.P., Zheng, W., Thomas, P.M., Kelleher, N.L., Bern, M., Schweitzer, M.H., (2017). Expansion for the *Brachylophosaurus canadensis* Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein, *J. Proteome Res.* 16, 920-932, <https://doi.org/10.1021/acs.jproteome.6b00873>.
- 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, 2603, <https://doi.org/10.7717/peerj.2603>.

Extinct species identification from Upper Pleistocene bone fragments not identifiable from their osteomorphological studies by proteomics analysis

- Scott, I., Yamauchi, M., Sricholpech, M., (2012). Lysine post-translational modifications of collagen, *Essays Biochem.* 52, 113-133, <https://doi.org/10.1042/bse0520113>.
- 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>.
- Surmik, D., Boczarowski, A., Balin, K., Dulski, M., Szade, J., Kremer, B., Pawlicki, R., (2016). Spectroscopic Studies on Organic Matter from Triassic Reptile Bones, Upper Silesia, Poland, *PLOS One* 11, 0151143, <https://doi.org/10.1371/journal.pone.0151143>.
- 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>.
- Veitschegger, K., Kolb, C., Amson, E., Scheyer, T.M., Sanchez-Villagra, M.R., (2018). Palaeohistology and life history evolution in cave bears, *Ursus spelaeus sensu lato*, *PLOS One* 13, 0206791, <https://doi.org/10.1371/journal.pone.0206791>.
- Vinciguerra, R., De Chiaro, A., Pucci, P., Marino, G., Birolo, L., (2016). Proteomic strategies for cultural heritage: From bones to paintings, *Microchem. J.* 126, 341-348, <https://doi.org/10.1016/j.microc.2015.12.024>.
- Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., (2012). The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013, *Nucleic Acids Res.* 41, D1063-D1069, <https://doi.org/10.1093/nar/gks1262>.
- Wadsworth, C., Buckley, M., (2014). Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone, *Rapid Commun. Mass Spectrom.* 28, 605-615, <https://doi.org/10.1002/rcm.6821>.
- Welker, F., (2018). Palaeoproteomics for human evolution studies, *Quat. Sci. Rev.* 190, 137-147, <https://doi.org/10.1016/j.quascirev.2018.04.033>.
- Welker, F., Collins, M.J., Thomas, J.A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S.T., Reguero, M., Gelfo, J.N., Kramarz, A., (2015). Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, *Nature* 522, 81, <https://doi.org/10.1038/nature14249>.
- Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., (2016). Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, *Proc. Natl. Acad. Sci.* 113, 11162-11167, <https://doi.org/10.1073/pnas.1605834113>.
- Welker, F., Ramos-Madrugal, J., Kuhlwilm, M., Liao, W., Gutenbrunner, P., de Manuel, M., Samodova, D., Mackie, M., Allentoft, M.E., Bacon, A.-M., (2019). Enamel proteome shows that Gigantopithecus was an early diverging pongine, *Nature* 576, 262-265, <https://doi.org/10.1038/s41586-019-1728-8>.
- 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, 3033, <https://doi.org/10.7717/peerj.3033>.
- Wu, J., Kohno, N., Mano, S., Fukumoto, Y., Tanabe, H., Hasegawa, M., Yonezawa, T., (2015). Phylogeographic and Demographic Analysis of the Asian Black Bear (*Ursus thibetanus*) Based on Mitochondrial DNA, *PLOS One* 10, e0136398, <https://doi.org/10.1371/journal.pone.0136398>.
- Young, M.F., Kerr, J.M., Ibaraki, K., Heegaard, A.-M., Robey, P.G., (1992). Structure, expression, and regulation of the major noncollagenous matrix proteins of bone, *Clin. Orthop. Relat. Res.*, 275-294, <http://www.ncbi.nlm.nih.gov/pubmed/1499220>.
- Zaitseva, O.V., Shandrenko, S.G., Veliky, M.M., (2015). Biochemical markers of bone collagen type I metabolism, *Ukr. Biochem. J.* 87, 21-32, <https://doi.org/10.15407/ubj87.01.021>.