Earth and Environmental Science Transactions of the Royal Society of Edinburgh, **114**, 83–93, 2023

Characterising the cave bear *Ursus spelaeus* Rosenmüller by ZooMS: a review of peptide mass fingerprinting markers

Ana GARCÍA-VÁZQUEZ[1], Ana Cristina PINTO-LLONA[2], Julià MAROTO[3], Trinidad TORRES[4] and Aurora GRANDAL-D'ANGLADE[5]*

[1] ArchaeoScience Division, Research Institute of the University of Bucharest (ICUB), University of Bucharest, 90-92 Sos. Panduri, 5th District, 50663, Bucharest, Romania
[2] Museo Arqueológico Regional, Plaza de las Bernardas s/n. 28801 Alcalá de Henares, Spain
[3] Institut de Recerca Històrica, Universitat de Girona, Àrea de Prehistòria, pl. Ferrater Mora 1, 17004 Girona, Catalonia, Spain
[4] Biomolecular Stratigraphy Laboratory, E.T.S.I. Minas, Universidad Politécnica de Madrid, C/Rios Rosas 21, 28003 Madrid, Spain
[5] Molecular Palaeontology Laboratory, Instituto Universitario de Xeoloxía, Universidade da Coruña, ESCI, Campus de Elviña, 15071 A Coruña, Spain
*Corresponding author. Email: aurora.grandal@udc.es

ABSTRACT: In the last decade, the identification of bone fragments by peptide mass fingerprinting or zooarchaeology by mass spectrometry is developing as a powerful tool in Quaternary palaeontology. The sequence of amino acids that make up the bone collagen molecule shows slight variations between taxa, which can be studied by mass spectrometry for taxonomic purposes. This requires reference databases that allow peptide identification. Although the cave bear (*Ursus spelaeus* Rosenmüller, 1794) is a common component in many European Pleistocene cave sites, no peptide fingerprint taxonomic study has paid special attention to this species up to now. For peptide markers in Ursidae, the most recent proposal is based on collagen obtained from a modern brown bear sample. In this work we attempt to cover this gap by studying bone collagen of cave and brown bear samples from different origins and different chronology, applying matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI TOF). We also performed an *in-silico* study of ursid bone collagen sequences published in databases. In our results we detected some discrepancies between the peptides obtained from both *in silico* and MALDI TOF analysis of fossil collagen and those published in the literature, in which we conclude that there are some misidentified peptides. The identification of skeletal remains by means of their peptide fingerprint is proving to be a powerful tool in palaeontology, which will bear greater fruit once the limitations of a technique that is in its initial stages have been overcome.

KEY WORDS: bone collagen, molecular palaeontology, Pleistocene, Ursidae.

Biology and palaeontology differ, among other things, in that the former studies living beings that perform all their vital functions and the latter studies vestiges of living beings that are currently integrated in the sedimentary record. However, there is an intermediate field relating to those beings that, although no longer alive, still retain some of their organic components unaltered. Apart from the exceptional preservation that occurs in environments with special conditions (freezing, mummification, etc.), the parts of animals that remain unaltered for the longest time after death are mainly mineralised tissues, such as bone or dentine.

Within this mineralised casing, biomolecules such as DNA or certain normally labile proteins can survive for thousands of years (Buckley & Collins 2011). Collagen is one such protein, and in recent decades it has been the subject of molecular palaeontology, the branch of palaeontology that studies the molecules contained in fossils in order to reconstruct various biological aspects of organisms from the past.

As the protein that it is, fossil collagen contains genetic information that in the last decade is being used for taxonomic purposes using the technique called zooarchaeology by mass spectrometry (ZooMS, Buckley *et al*. 2009). In this paper we will review the structure and composition of collagen and how it is applied to the identification of Pleistocene and Holocene European ursid remains.

1. Bone collagen as a subject of palaeontological study

1.1. On the bone collagen structure and composition

Bone is a living tissue made up of cells and extracellular matrix. It also has blood and lymphatics vessels, and nerve endings. The

 doi:10.1017/S1755691023000038

cells are found in spaces called lacunae and are responsible for secreting the components of bone, maintaining and remodelling it (Davies & Hosseini 2000). The bone extracellular matrix is made up of two main components: the mineral fraction; and the organic matrix. The mineral fraction is mainly calcium phosphate or hydroxyapatite. The organic matrix, which accounts for one-third of the bone weight, consists mainly of proteins, including collagen (90%) and other non-collagenous proteins (3 to 5%) to which mineral crystals are bound (Davies & Hosseini 2000). Dentine is similar in composition to the extracellular matrix of bone but lacks cells and does not undergo remodelling.

Type 1 collagen is the major insoluble fibrous protein of the organic matrix of bone or dentine (Henriksen & Karsdal 2019). Like other proteins, collagen has a primary structure determined by the amino acid sequence. It is based on a repeat of the sequence glycine–X–Y, where X and Y can be any of the other amino acids except cysteine or tryptophan (Eastoe 1955; Robinson & Rudd 1974). Each molecule, or α-strand, acquires a three-dimensional secondary structure in the form of a left-handed helix. Individual strands are unstable and must aggregate for stability. Mature collagen or tropocollagen acquires a quaternary structure formed by three α-strands, twisted together into a right-handed triple helix. In tetrapods, the three strands of the triple helix of type 1 collagen are not the same: there are two α1-strands; and one α2-strand (Ricard-Blum 2011). Collagen synthesis takes place mainly in bone tissue cells, where α1 and α2 chains are synthesised separately (Henriksen & Karsdal 2019). Both types of chains gradually mature losing the signal peptide and both terminal propeptides, leaving the helical region formed by more than 1000 residues (amino acids), flanked by two shorter telopeptides. The helical region of each chain undergoes post-translational modifications, which will eventually allow the three chains to wrap around each other to form the characteristic triple helix (Shoulders & Raines 2009). The main modification is the hydroxylation of the amino acids, proline (Pro, P) and lysine (Lys, K). Hydroxylation is a chemical reaction in which a hydroxyl (OH) group is introduced to replace an H atom. Hydroxylation of proline and lysine can occur when they are in the third position of the sequence glycine–X–Y, and allows stable triple helices to be formed (Kuhn 1987). In type 1 collagen, approximately 50% of proline residues are hydroxylated, whereas this modification is more variable in lysine, between 15 and 90% of residues (Yamauchi & Sricholpech 2012). Finally, each triple-helix associates into a right-handed super-super-coil referred to as the collagen microfibril. Each microfibril is interdigitated with its neighbouring microfibrils, forming the collagen fibre.

1.2. Fossil bones and fossil collagen

The bones of a present-day vertebrate contain on average 22% collagen (Crockett *et al.* 2011). In fossils, the proportion decreases, depending on the age of the remains and the pH, humidity and temperature conditions to which they are subjected. Diagenetic degradation of bone can follow several pathways. The dissolution of the mineral fraction of the bone, due to the acidity of the soil, leads to the accelerated loss of collagen as it becomes accessible to microbial attack. Chemical degradation of collagen is dependent on moisture and temperature. It has been estimated that, under current European climatic conditions and with no other factors affecting the bone, it could survive for over a million years (Buckley & Collins 2011). Such ideal conditions do not usually occur in nature; even so, it is possible to recover sufficiently preserved collagen in bone remains tens of thousands or even several hundred thousand years old (Buckley & Collins, 2011; Britton *et al.* 2012). In very rare cases, traces of collagen were found in dinosaur bones about one hundred million years old (Schweitzer *et al.* 2009; Lee *et al.* 2017). Experimental studies showed that the composition of collagen remains virtually unchanged until only 1% of the initial amount remains in the bone (Dobberstein *et al.* 2009).

The long preservation of collagen in bones and teeth allows biological and evolutionary data to be obtained from organisms that ceased to live long ago. Biomolecular analyses of collagen are increasingly used to reconstruct past life, whether in humans or animals. Stable isotope analyses of fossil bone collagen began to be applied in the last quarter of the 20th century, although initially only in an exploratory manner. However, their use has grown exponentially to the present day (Katzenberg & Waters-Rist 2019). As for taxonomic identification by mass spectrometry or peptide fingerprinting, it is a very recently implemented technique and is still under development. In zooarchaeology it has been specifically termed ZooMS (Buckley *et al.* 2009).

1.3. The ZooMS technique

The identification of bone remains by means of their collagen peptide fingerprint, or ZooMS, is proving to be a powerful tool in palaeontology. In Pleistocene sites, and even more so if they are of anthropogenic origin, the taxonomic identification of faunal remains is not always easy due to the high degree of fragmentation that bones usually present. To identify these small remains, one possibility would be to sequence their DNA. However, this is an expensive and laborious technique, which does not always give good results because the DNA of ancient remains is usually degraded and contaminated.

Another possibility is the sequencing of bone proteins, for example, collagen. The advantage over the study of DNA is that collagen is much more abundant in bone remains than DNA, is better preserved and is easier to extract. The disadvantage is that collagen is less specific. Within each major taxonomic group (e.g., mammals), the proteins that perform the same function are usually very similar, although they have small differences related to the evolution of the different lineages separately and the accumulation of mutations over time. Some very specific proteins, such as collagen, cannot accumulate too many mutations because they would lose their structure and function. Even so, small differences in amino acid sequence occur between collagen molecules of different taxa that allow differentiation of these taxa (Buckley 2018). In order to identify the taxon from which the bone remnant from which the collagen is obtained comes, it would be necessary to sequence the complete protein, that is, to identify the entire sequence of amino acids of which it is composed. This is a complex technique and only represents a small advantage over DNA sequencing, which is the greater ease of extraction of collagen compared to DNA, due to its abundance.

Peptide mass fingerprinting (PMF) (James *et al.* 1993; Pappin *et al.* 1993) is a technique for protein identification that does not require sequencing and is based simply on the differences in the molecular weight of the various peptides. Since each amino acid has a specific molecular weight, due to its chemical composition, small differences in the amino acid sequence will produce proteins of different molecular weight. If we consider the entire collagen molecule, in which each α-strand contains more than 1000 amino acids, the difference in molecular weight will be very small and it is not possible to know in which position the substitution (s) are located. The basis of the PMF technique consists of cleaving the protein at specific sites by enzymatic digestion. This yields a series of peptides of different sizes, each of which has a specific amino acid sequence and thus a characteristic mass. In this way it is easier to know at which points in the sequence (in which of the peptides) the substitutions are found.

Trypsin is an enzyme produced in the pancreas and acts in the digestion of proteins, breaking them into smaller fragments called peptides. This cleavage is specific, as it only acts by breaking the peptide bonds at the C-terminal end of the amino acids,

https://doi.org/10.1017/S1755691023000038 Published online by Cambridge University Press

Table 1 List of peptide markers identified so far for brown bear and polar bear.

Peptide marker identification		*m/z* for *Ursus*			
Welker *et al.* (2016)	Brown *et al.* (2021a)	*Ursus* (Buckley & Collins, 2011)	*Ursus arctos* (Kirby *et al.* 2013)	*Ursus arctos* (Welker *et al.* 2016)	*Ursus* (Buckley *et al.* 2017a)
P1	COL1α1 508–519	–	1105	1105.6	–
A	COL1α2 978–990	1233.7	1233	1217.7	1217.7
A'	COL1α2 978–990	–	–	1233.7	1233.7
B	COL1α2 484–498	1453.7	1453	1453.7	1453.7
C	COL1α2 502–519	1566.8	1566	1566.8	1566.8
P2	COL1α2 292–309	–		1609,8	–
D	COL1α2 793–816	1263,1	1263	2163.1	2163.1
E	COL1α2 454–483	–	–	(?)	–
F	COL1α1 586–618	2853.5 + 2869.5	2853	2853.4	2853.4
F'	COL1α1 586–618	–	–	2869.4	–
G	COL1α2 757–789	2957.4	2957	2957.5	2957.5
G'	COL1α2 757–789	–	–	2973.5	–

The markers identified as A', F' and G' are the same as their namesakes, but with an extra hydroxylation that adds 16 Da to the peptide.

lysine (Lys, K) and arginine (Arg, R), except when the following residue is a proline (Olsen *et al.* 2004). Digestion of a protein with trypsin *in vitro* produces the same cleavage effect on the protein under study. This results in a series of peptides (which, because they are obtained by trypsin, are called tryptic peptides) that will always end in a lysine or an arginine.

The substitution of a single amino acid produces a peptide of different molecular weight. These small differences cause homologous peptides in different taxa to have different masses. By measuring these masses, it is possible to recognise the different taxa without the need for sequencing or reading of the amino acid sequence. The application of ZooMS to identify bone proteins from ancient remains began to be developed at the beginning of the 21st century (Ostrom *et al.* 2000) and is still under development. Given the highly conserved amino acid sequence of collagen, most of the peptides obtained by tryptic digestion are identical across taxa and only a few of them are useful for differentiation at the taxonomic level. Collagen type 1 alpha 1 (col1α1) is more conservative among taxa (Buckley 2018) and only two of its peptides are used as markers for taxonomic purposes. The remaining peptide markers are found on alpha chain 2 (col1α2).

The first studies using ZooMS were focused on the identification of large terrestrial mammals (Buckley & Collins 2011; Buckley & Kansa 2011; Buckley *et al.* 2017a, among others), but their use soon expanded to marine mammals (Kirby *et al.* 2013; Buckley *et al.* 2014), micromammals (Buckley *et al.* 2016) or marsupials (Buckley *et al.* 2017b; Peters *et al.* 2021). Significant progress is also being made in the identification of fish (Richter *et al.* 2011; Harvey *et al.* 2018), bird bone (Horn *et al.* 2019) and eggshell (Stewart *et al.* 2013; Presslee *et al.* 2017), amphibians (Buckley & Cheylan 2020) or sea turtles (Harvey *et al.* 2019), among others, which reveals the great potential of the use of ZooMS.

1.4. Identification of Ursids by ZooMS

Although the cave bear (*Ursus spelaeus sensu lato*) is a common component in many European Pleistocene cave sites, no peptide fingerprint taxonomy study has paid special attention to this species up to now. The first work devoted to the application of peptide mass fingerprinting to fossil mammals (Buckley *et al.* 2009) does not include any ursid specimens. Subsequently, bear peptide markers were offered in some works (Buckley & Collins 2011; Kirby *et al.* 2013; Welker *et al.* 2016; Buckley *et al.* 2017a). However, the recent implementation of this technique and the fact that it is still under development causes some disparity in the definition of peptides, either in their validity as markers, or in their position in the molecule, or simply in their nomenclature (Richter *et al.* 2022). For peptide markers in Ursidae, the most recent proposal is that of Welker *et al.* (2016), which is based on collagen obtained from a modern brown bear sample. The *m/z* values of the marker peptides (Table 1) do not differ from those proposed in previous or subsequent works.

Initially the peptide markers were identified as correlative letters of the alphabet (Buckley *et al.* 2009). Subsequently, a system was adopted that identifies peptides by indicating which chain they come from (α1 or α2) and their order in the molecule (Buckley *et al.* 2009). A recent proposal for standardisation of peptide nomenclature (Brown *et al.* 2021a) in addition to indicating the chain, identifies each peptide by the position of the amino acid with which it begins and ends starting from the beginning of the helical region, which facilitates the task of identifying the peptides.

Since none of the works identifying peptide markers included specifically cave bear collagen sequences, in this work we will attempt to cover this gap by studying cave bear samples from different origins and different chronology. The purpose of this study is twofold: firstly, we will check whether the marker peptides proposed in the literature for brown bears coincide with those obtained in cave bears; and secondly, we will try to find out if there is any difference between both cave and brown bears from distant geographical regions or of different chronology. For our purpose, we will rely on the direct study of collagen samples, but also on the *in-silico* study of the sequences available in protein databases. This allows, through the use of bioinformatics tools, to obtain the theoretical tryptic peptide spectrum and compare it with those obtained from bone samples.

2. Material and methods

2.1. *In silico* study of the sequences of ursid bone collagen

One way to know the sequence of amino acids that make up the a and b chains of ursid collagen is to resort to the databases available in UniProt Knowledgebase (UniProtKB). This is a central hub of protein knowledge by providing a unified view of protein sequence and functional information, made freely available by The UniProt Consortium in https://www.uniprot.org/uniprot/ (Magrane 2011).

The UniProtKB consists of two sections: UniProtKB/Swiss-Prot; and UniProtKB/TrEMBL. UniProtKB/Swiss-Prot is manually curated, so that the information in each entry is annotated and reviewed by a curator. The records in UniProtKB/TrEMBL are automatically generated and the records await full manual annotation. This means that not all the entries are fully reliable.

https://doi.org/10.1017/S1755691023000038 Published online by Cambridge University Press

Table 2 Sequences of col1α1 and col1α2 in UniProtKB used in this work.

Taxon		col1α1 entry	col1α2 entry
Brown bear	*Ursus arctos*	A0A3Q7X3Q3	A0A3Q7VKW
Polar bear	*Ursus maritimus*	A0A384BX56	A0A384BPF6
Dog	*Canis lupus familiaris*	Q9XSJ7	O46392
Cat	*Felis sylvestris catus*	M3W2F5	M3WVN3
Boreal lynx	*Lynx canadensis*	A0A667J4W5	A0A667GAY3
Tiger	*Panthera tigris*	A0A8C9M8Y4	A0A8C9K5D1
Lion	*Panthera leo*	A0A8C8Y4U8	A0A8C8XMR9
Leopard	*Panthera pardus*	A0A6P4TFM8	A0A6P4VEL6
Ferret	*Mustela putorius furo*	M3YVG8	M3XR96
Human	*Homo sapiens*	P02452	P08123
Cow	*Bos taurus*	P02453	P02465
Sheep	*Ovis aries*	W5P481	W5NTT7
Horse	*Equus caballus*	F6SSG3	F6RTI8

In the case of the Ursidae, none of the sequences of collagen 1 are manually revised. To choose the most reliable entries, we used the sequence alignment function and chose those sequences that share the most common positions.

For col1α1 there are three sequences, one from *Ursus maritimus*, the polar bear (A0A384BX56 in UniprotBK) and two from *Ursus arctos horribilis* (the North American grizzly bear), of which only one (A0A3Q7X3Q3) preserves the complete helical region (that which constitutes the mature collagen fibrils). The alignment shows that, although there are some differences at the ends of the molecules, the helical region, which is used to identify marker peptides, is almost completely coincident (only two substitutions in a molecule of more than 1000 residues).

For col1α2, there are 10 sequences from *U. maritimus* and only one from *U, arctos horribilis*. Only one of the *U. maritimus* sequences (A0A384BPF6_URSMA) is almost complete. Performing an alignment of this entry with the *U. arctos horribilis* one, the coincidence in the amino acid sequence is 100% in the helical region. Since the sequence is identical in both species, we consider it to be valid. In addition, for comparative purposes, we have used the sequences of other carnivore species (Table 2). Of these, the dog is the only one that is manually checked. For the carnivores we added punctually the comparison with the collagen sequences of human, cow, sheep and horse.

The sequences obtained were analysed using the Peptide Mass tool available on Expasy, (Swiss Bioinformatics Resource Portal, https://www.expasy.org/). This tool allows performing a tryptic digestion simulation and obtaining the peptide spectrum of each type of collagen, with its amino acid sequence and the theoretical *m/z* value. To this value it will be necessary to add the difference for each possible hydroxylation of the P or K residues. In addition, the deamidation of glutamine (Gln, Q), a frequent alteration in ancient collagen (Van Doom *et al*. 2012; Wilson *et al*. 2012), would add +0.984 Da (practically one unit) for each altered glutamine (Robinson & Rudd 1974).

2.2. Cave and brown bear bone samples

For this study we have selected 20 samples of cave bear, *U. spelaeus* Rosenmüller, 1794, identified morphologically and, in most cases, genetically (Table 3). The samples come from several sites in the Iberian Peninsula, Austria, Italy, Slovenia and Russia (Fig. 1) and cover most of the cave bear varieties described according to their mitochondrial lineages: *U. spelaeus*; *Ursus ingressus*; *Ursus ladinicus*; *Ursus rossicus*; and *Ursus kanivetz* (Barlow *et al*. 2021). Direct carbon-14 dating is available for five of them, while the ages of the remaining samples are known from their stratigraphic position.

Additionally, we included in the study 10 samples of brown bear (*Ursus arctos* Linnaeus, 1758), all identified morphologically and most of them also by their mitochondrial DNA. The samples come from the Cantabrian region in the Iberian Peninsula. Two are of Pleistocene age, seven are Holocene and the last one is a modern specimen, deceased in 2015 in the vicinity of the town of Belmonte de Miranda (Asturias, Spain).

2.3 Pretreatment of the samples and extraction of bone collagen

A sample of approximately 1 g was cut from each bone with a hand tool equipped with a diamond disc. Cancellous tissue and superficial concretions, if present, were mechanically removed. The fragment was repeatedly rinsed in an ultrasonic bath, successively in deionised water and acetone (a minimum of five rinses in acetone and six rinses in water, or more if necessary, until no turbidity was observed) and then left to dry in glass Petri dishes that protect it from dust and other possible contaminations, at room temperature, for at least 48 h.

The collagen extraction protocol follows a modified method of Longin (1971) described in Bocherens *et al*. (1997), with further modifications implemented in the Laboratory of Molecular Palaeontology of the University Institute of Geology, University of A Coruña (Spain), where the treatment was carried out. Our purification protocol is based on successive filtrations, which eliminate collagen fragments and retain only the large collagen strands. For each specimen studied, bone fragments (about 500 mg) were manually ground with agate mortar and pestle. The bone powder was sieved to obtain the fraction less than or equal to 0.5 mm. The use of powdered bone shortens the demineralisation time and therefore reduces the possibility of collagen degradation.

From each sample, a portion of between 250 and 300 mg of bone powder was taken, demineralised in about 30 mL of 1 M hydrochloric acid for 20 min, washed in deionised water until reaching a neutral pH and filtered through nitrocellulose filters (Sartorius Stedim®) of 5 µ pore size. The solid residue was incubated for 21 h at room temperature in 30 mL of 0.125 M sodium hydroxide to remove possible organic contaminants, such as fats or humic acids. After washing at neutral pH and further purification by filtration, the solid fraction containing collagen was solubilised in 20 mL of 0.1 M hydrochloric acid for 17 h at 90°, filtered a third time to remove insoluble mineral particles, frozen at −80 °C and freeze-dried for the analysis.

2.4. Collagen analysis

For peptide fingerprinting or ZooMS analysis, an aliquot of the collagen isolated from each bone was digested with trypsin, which breaks the molecular bonds between specific amino acids (after a lysine, K or an arginine, R, if not followed by proline, P). Thus, a set of peptides of different mass and charge (*m/z*) was obtained, identified by matrix-assisted laser desorption/ionisation, time-of-flight (MALDI-TOF). This analysis was performed at the Mass Spectrometry and Proteomics Unit of the Research and Technological Development Support Infrastructures Network, University of Santiago de Compostela (Spain).

For each sample, an aliquot of 1 to 5 mg of lyophilised collagen was dissolved in ammonium hydrogen carbonate buffer. After addition of Promega Trypsin Gold, mass spectrometry grade, the samples were digested at 37 °C overnight. The sample solution was mixed with a matrix of α-cyano-4-hydroxycinnamic acid by applying 1 µl of the mixture onto the MALDI plate in a Bruker Ultraflex® III MALDI-TOF/TOF mass spectrometer equipped with a smart beam laser. The principle of this type of analyser is simple. Once the collagen-matrix mixture is introduced into the plate, laser shots cause the mixture to be gently

https://doi.org/10.1017/S1755691023000038 Published online by Cambridge University Press

Table 3 Data on the bone samples used in this study: location, chronology, taxonomy and collagen quality parameters. %Ccol: percentage of C in collagen. %Ncol, percentage of N in collagen. C:Nat, carbon-to-nitrogen atomic weight ratio in collagen.

Region	Locality	Cave	Sample	Taxon	Chronology	Reference	% Ncol	% Ccol	C: Nat	% col
Cantabrian	Galicia, IP	Eirós	E-3-89	*Ursus spelaeus* (1)	44,000 ± 1900	De Lombera (2020)	13.3	33.6	3.2	11.7
Cantabrian	Galicia, IP	Eirós	E-VS-55	*U. spelaeus* (1)	25,592 ± 602	Pérez Rama *et al.* (2011)	13.2	36.0	3.2	8.8
Cantabrian	Galicia, IP	Liñares	LIN-E-234	*U. spelaeus* (1)	>40	Pérez Rama *et al.* (2011)	8.9	24.0	3.1	21.9
Cantabrian	Galicia, IP	Liñares	VL-L-549	*U. spelaeus* (1)	>40	Pérez Rama *et al.* (2011)	9.1	24.7	3.1	7.1
Cantabrian	Galicia, IP	Liñares	LIN-1009	*U. spelaeus* (1)	>40,000	Pérez Rama *et al.* (2011)	15.8	42.5	3.2	10.5
Pyrennees	Navarra, IP	Amutxate	AX-1069	*U. spelaeus* (1)	39–48*	Torres *et al.* (2007)	12.4	33.3	3.1	7.5
Pyrennees	Navarra, IP	Amutxate	AX-4119	*U. spelaeus* (1)	39–48*	Torres *et al.* (2007)	10.0	27.5	3.2	3.7
Pyrennees	Catalonia, IP	Ermitons	ERM404	*U. spelaeus* (?)	≈30–35	Maroto (1993)	12.1	33.4	3.2	12.2
Pyrennees	Catalonia, IP	Ermitons	ERM405	*U. spelaeus* (?)	≈30–35	Maroto (1993)	11.0	30.3	3.2	11.0
Apuan Alps	Sienna, Italy	Chiostraccio	CHIOA2	*Ursus ingressus* (?)	23,930 ± 100	Martini *et al.* (2014)	11.0	31.0	3.3	10.3
Alps	Trento, Italy	Conturines	CO2	*Ursus ladinicus* (2)	40 to >49	Döppes *et al.* (2019)	10.7	27.4	3.0	8.6
Alps	Austria	Liegl loch	LL2	*U. ingressus* (2)	28,130 ± 600	Fernández *et al.* (2001)	9.8	25.5	3.0	6.8
Alps	Austria	Liegl loch	LL3	*U. ingressus* (2)	25–30	Pacher & Stuart (2009)	10.8	27.2	2.9	1.8
Alps	Austria	Schwabenreith	SW3	*Ursus eremus* (2)	35 to >49	Döppes *et al.* (2019)	13.1	33.3	3.0	27.9
Karawanks	Slovenia	Potocka Zijalka	PZ2	*U. ingressus* (2)	27–36	Pacher & Stuart (2009)	11.0	30.9	3.0	10.5
Urals	Russia	Kizel	KIZ1	*Ursus rossicus* (2)	≈32–36	Pacher & Stuart (2009)	15.8	43.4	3.2	17.3
Urals	Russia	Kizel	KIZ4	*U. rossicus* (2)	≈32–36	Pacher & Stuart (2009)	15.4	41.9	3.2	12.3
Urals	Russia	Medvezhiya	MED1	*Ursus kanivetz* (2)	≈42–45	Barlow *et al.* (2021)	15.9	42.3	3.1	11.4
Urals	Russia	Medvezhiya	ZIN RAS 34756	*U. kanivetz* (2)	41,940 ± 500	Grandal-d'Anglade *et al.* (2019)	16.1	42.9	3.1	13.6
Urals	Russia	Medvezhiya	MED3	*U. kanivetz* (2)	≈42–45	Barlow *et al.* (2021)	15.4	41.1	3.1	11.7
Cantabrian	Galicia, IP	Arcoia	ARLU-39	*Ursus arctos* (3*)	>40,000	García-Vázquez *et al.* (2011)	7.4	19.5	3.1	–
Cantabrian	Galicia, IP	Arcoia	ARLU-42	*U. arctos* (3)	31,710 ± 720	García-Vázquez *et al.* (2011)	14.7	39.7	3.1	7.5
Cantabrian	Galicia, IP	Pena Paleira	SIPA-61	*U. arctos* (1,3)	7201 ± 46	García-Vázquez *et al.* (2011)	13.3	34.9	3.1	–
Cantabrian	Galicia, IP	Pena Paleira	SIPA-215	*U. arctos* (4)	Holocene	Unpublished	13.9	38.1	3.2	9.8
Cantabrian	Galicia, IP	Purruñal	Pur-Lu-9	*U. arctos* (1,3)	7815 ± 80	García-Vázquez *et al.* (2011)	14.8	40.5	3.2	10.9
Cantabrian	Asturias, IP	Pozu Toneyo	RT-001	*U. arctos* (1,3)	8800 ± 40	García-Vázquez *et al.* (2015)	14.6	39.5	3.2	4.0
Cantabrian	Asturias, IP	Pozu la Cigacha	CGLL-051	*U. arctos* (1,3)	6750 ± 40	García-Vázquez *et al.* (2015)	13.9	38.1	3.2	–
Cantabrian	Asturias, IP	Sima Osos Somiedo	SH5-98-S28-095	*U. arctos* (1,3)	8990 ± 50	Pinto Llona *et al.* (2005)	15.3	41.3	3.2	–
Cantabrian	Asturias, IP	Belmonte de Miranda	BEL-1	*U. arctos*	Modern (†2015)	García-Vázquez *et al.* (2018)	12.9	34.4	3.1	30.8
Cantabrian	Cantabria, IP	Unknown	GP-1	*U. arctos* (3)	2410 ± 30	García-Vázquez *et al.* (2015)	15.5	41.7	3.1	3.7

The taxon column refers, for cave bears, to the genetic identification of the cave bear variety according to its mitochondrial DNA: (1) according to González-Fortes *et al.* (2016); (2) according to Barlow *et al.* (2021); (3) González-Fortes *et al.* (2017); (3*) González-Fortes *et al.* (2017) but no DNA was recovered; and (4) no direct DNA study. (?) indicates the lack of genetic study in the site. The age of the samples is given in carbon-14 dating ages BP and their error when the sample is directly dated; in ka BP when it is a stratigraphic age (obtained by dating other coeval bones). (*) dates obtained by amino acid racemisation. Abbreviations: IP = Iberian Peninsula.

ionised. Next, an extraction voltage induces the mobilisation of all peptides simultaneously. These will pass through an accelerating electrostatic field, acquiring a high kinetic energy that propels them in the direction of the flight tube facing the detector. The travel time of the flight tube length will be directly proportional to the mass to charge ratio (*m*/*z*) of the ionised peptides. The spectrum produced is compared with published reference spectra (Welker *et al.* 2016, the most complete and recent database) and with those obtained from *in silico* tryptic digestion, to identify the peptide markers and their *m*/*z* values.

3. Results

Collagen had been extracted from all the samples included in this work for stable isotope studies. Some results have been published (Pérez-Rama *et al.* 2011; García-Vázquez *et al.* 2018; Grandal-d'Anglade *et al.* 2019), while others are currently in preparation. Collagen from all samples yielded good results in terms of the usual quality criteria: yield; % carbon (C) and (N); and C:N atomic ratio (DeNiro 1985; Ambrose 1990; Van Klinken 1999).

The results obtained from tryptic digestion and MALDI-TOF analysis are shown in Table 4, where only the marker peptides are listed following Welker *et al.* (2016). The full spectra can be seen in the Online Supplementary Material (Table S1) available at https://doi.org/10.1017/S1755691023000038. Despite the absence of some markers in some samples, this result allows us to identify all specimens as members of the genus *Ursus*.

4. Discussion

4.1. Identification of peptides in Ursidae

The peptide spectra obtained from the collagen analysis of 30 ursid specimens show in general a good coincidence in some of the markers, such as P1, B, D and G, while in others the presence is variable, or even does not appear in any of the samples.

The absence of peptide A or COL1α2 978–990 is not unusual, as we found it mainly as the variety with an extra hydroxylation (A'). In contrast, the peptide COL1α2 767–799 is preferably found in its version with one less hydroxylation than all the others (G instead of G'). In most cases it shows one extra unit, the result of the deamidation of the single glutamine (Q) residue it contains.

The absence of some of the peptides in some of the samples may be due to the collagen molecule being broken during the diagenetic phase, so that digestion with trypsin produces peptides of a smaller size than expected. This tends to occur if the bones are badly damaged and is most noticeable in the larger peptides (Buckley *et al.* 2011). However, the low presence or even

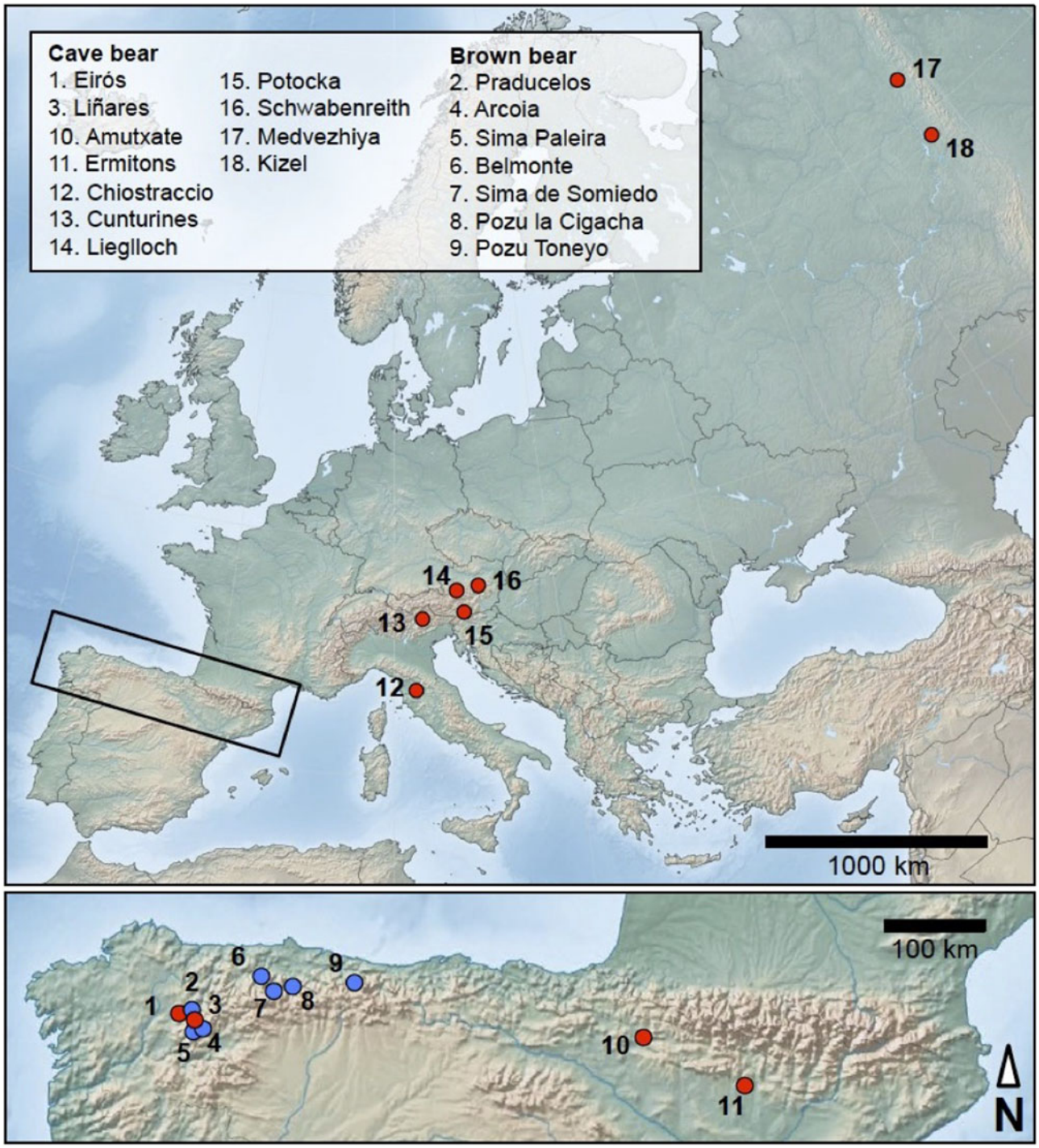

Figure 1 Map showing the location of the sites from which the studied samples originated. Below, an enlargement of the inset of the map above.

complete absence of some peptides such as P2, E or F, which we will discuss in more detail below, is noteworthy.

4.1.1. P2 peptide, COL1α2 292–309. This peptide was initially proposed as useful for cetacean identification (Buckley *et al*. 2014). In terrestrial mammals, and according to previous studies, it is identified by a peak of *m/z* 1609.8, common to all canids, felids and mustelids for which there are ZooMS data (Welker *et al*. 2016).

However, our *in-silico* study on bear and other carnivores' col1α2 sequences shows at that position the sequence GPNGEAGSAGPSGPPGLR whose *m/z* is 1577.7. It contains a proline (P) susceptible to hydroxylation (before glycine, G), so the peptide could reach a *m/z* of 1593.7. The same sequence is found in felids and canids. In addition to this discrepancy, the only taxonomic study that includes a significant amount of ursid samples, from Denisova Cave (Brown *et al*. 2021b) does not identify this peptide in any of the 175 samples identified as ursids. Nor does it appear in any of the carnivores of that site.

In our cave bear set, a 1609 peak does not appear in any of the 20 cases, but a 1592 mass peptide appear in 18 of them, and 1593 in another one. The same is true for brown bears, where the peak 1609 is absent in all the samples, with a peak at 1592 in seven of the samples and one with 1593. This peptide could correspond to COL1α2 292–309, but the systematic difference of one unit less than the theoretical mass in most of the samples does not allow us to state this before applying a technique capable of identifying each residue, and not just the mass of the whole peptide. This lower-than-expected value may be due to cave bears showing low δ^{15}N and δ^{13}C bone collagen values (Grandal *et al*. 2019) due to their plant-based diet. If the isotopic ratios of the samples

https://doi.org/10.1017/S1755691023000038 Published online by Cambridge University Press

Table 4 Results of the identification of peptide markers in the analysed samples.

Site	Sample	α1 508–519 (P1)	α2 978–990 (A)	+16 (A')	α2 484–498 (B)	α2 502–519 (C)	α2 292–309 (P2)	α2 793–816 (D)	α2 454–483 (E)	α1 586–618 (F)	+16 (F')	α2 757–789 (G)	+16 (G')
Brown bear	Welker *et al.* (2016)	1105	1217	1233	1453	1566	1609	2163	(?) 2808	2853	2869	2957	2973
Eirós	E-3-89	1105	x	1233	1453	1566	x	2163	2808	2853	2869	2957	x
Eirós	E-VS-55	1105	x	1233	1453	1566	x	2163	x	x	x	2958	x
Liñares	LIN-E-234	x	x	x	1453	x	x	2163	x	x	x	2958	x
Liñares	VL-L-549	1105	x	x	1453	1565	x	2163	x	x	x	2958	x
Liñares	LIN-1009	1105	x	x	1453	1565	x	2163	x	x	x	2958	x
Amutxate	AX1069	1105	x	1233	1453	x	x	2163	x	x	x	2957	x
Amutxate	AX4119	1105	x	1233	1453	1566	x	2163	x	2853	x	2958	x
Ermitons	ERM404	1105	x	x	1453	1566	x	2163	2808	2854	x	2958	x
Ermitons	ERM405	1105	x	x	1453	1566	x	2163	x	2853	x	2958	x
Chiostraccio	CHIOA2	1105	x	1233	1453	x	x	2163	x	x	x	2958	x
Conturines	CO2	1105	x	1233	1453	x	x	2163	x	x	x	2957	x
Lieglloch	LL2	1105	x	1233	x	x	x	2163	x	2853	x	2958	x
Lieglloch	LL1	1105	x	1233	1453	1565	x	2163	x	x	x	2958	x
Schwabenreith	SW3	1105	x	1233	1453	x	x	2163	x	x	x	2958	x
Potocka Zijalka	PZ2	1105	x	1233	1453	1566	x	2163	2808	x	x	2957	x
Kizel	KIZ1	1105	x	x	1453	1565	x	2163	x	x	x	2958	x
Kizel	KIZ4	1105	x	x	1453	1566	x	2163	x	x	x	2957	x
Medvezhiya	MED1	1105	x	1233	1453	x	x	2163	2808	x	x	2957	x
Medvezhiya	MED2	1105	x	x	1453	1565	x	2163	x	x	x	2957	x
Medvezhiya	MED3	1105	x	x	1453	1566	x	2163	x	x	x	2958	x
Arcoia	ARLU39	1105	x	x	1453	x	x	2163	x	x	x	2958	x
Arcoia	ARLU42	x	x	x	1453	x	x	2163	x	2853	x	2958	x
Paleira	SIPA61	1105	x	1233	1453	1566	x	2163	x	x	x	2958	x
Paleira	SIPA215	1105	x	1233	1453	1566	x	2163	x	2853	x	2958	x
Purruñal	PURLU9	1105	x	1233	1453	1566	x	2163	x	x	x	2957	2974
Pozo Toneyo	RT01	1105	x	x	1453	1565	x	2163	x	x	x	2957	x
Pozu La Cigacha	CGLL51	1105	x	1233	1453	1566	x	2163	x	x	x	2957	x
Sima Somiedo	S1SH5	1105	x	1233	1453	1566	x	2163	x	x	x	2957	x
Unknown	GP-1	1105	x	x	1453	x	x	2163	x	x	x	2957	x
Belmonte	BEL1	1105	x	1233	1453	1566	x	2163	x	x		2957	x

are different from those calculated *in silico*, based on a homogeneous isotopic mixture, it would not be impossible to obtain a slightly lower *m/z* value. It could be argued that this reason is not valid for the more omnivorous brown bear, but all our samples come from the Cantabrian area, where brown bears showed a diet mainly based on plant foods (García-Vázquez *et al.* 2018). We also found this possible effect in the peptide Col1α2 502–519 (C) where five of the cave bears also show one unit less than expected, while there is only one in the brown bear set. In any case, all results point to the fact that this peptide was misidentified in carnivores, or at least in ursids, in the databases published so far.

4.1.2. Peptide E, Col1α2 454–483. The validity of this peptide as a taxonomic marker was initially proposed by Buckley *et al.* (2011) but was later excluded from further studies by those authors since it was not present in many ancient samples (Buckley *et al.* 2017a). Recent studies, however, keep considering it among the useful markers. Welker *et al.* (2016) assign this peptide a *m/z* of 2808 for the polar bear, whereas for brown bear and American black bear it is unresolved. These authors rely on sequences obtained from modern samples, but their sequences are not complete (Welker *et al.* 2016, supplement). The antiquity, however, does not seem to be the reason for the absence of this peptide in our samples, as discussed below.

In the analysis of the Col1α2 sequences of UniProtKB, both the brown bear (A0A3Q7VKW6) and the polar bear (A0A384BPF6) show that positions 454 to 483 are occupied by the sequence GEQGPAGPPGFQGLPGPAGT AGEAGKPGER, with *m/z* = 2744.3. The sequence contains three proline residues susceptible to hydroxylation and two glutamine (Q) residues susceptible to deamidation. This offers a variety of possibilities as to the final *m/z* of the peptide. With all three prolines hydroxylated, the value would be 2792. If we add the possibility of Q deamidation, the final value could be 2793 or 2794.

Brown *et al.* (2021b), in the Denisova Cave samples, record peaks of *m/z* 2792 in 10 samples of the 175 identified as ursids, and none with the value 2808. In our 20 cave bear samples, 15 yield a peak at 2793 and five at 2794. Additionally, only four of them show a peptide 2808. In the case of brown bears, seven of them show a peak at 2793, and none at 2808, including the present-day brown bear sample.

Based on our results and the absence of peptide 2808 in the Denisova spectra, we propose that the peptide value in cave bears and by extension in Ursidae is 2792, which could be 2793 or 2794 considering the possible deamidation of the Q residues in the ancient samples.

4.1.3. Pepide F or COL1α1 586–618. It is located in the α1 chain of bone collagen. In ursids it is identified by a peak at *m/z* 2853.4. However, *in silico* digestion of Col1α1 from any of the mammals we used in this study does not yield this peptide intact, but rather two contiguous peptides, at positions 586 to 603 and 604 to 618.

The first peptide, which we will call here F1 for short, is composed of the sequence GLTGPIGPPGPAGAPGDK which is invariant not only in ursids or carnivores, but in all the taxa reviewed. Its *m/z* is 1558.8, but as it presents two prolines in position suitable for hydroxylation, the *m/z* values could be increased to 1574.8 with one hydroxylation, and 1590.8 with

https://doi.org/10.1017/S1755691023000038 Published online by Cambridge University Press

Table 5 Alternative values for some peptide markers identified in the *in-silico* analysis of α1 and α2 collagen sequences and their presence in the analysed samples.

Taxon	Sample	α2 292–309 (P2)	α2 454–483 (E)	α1 586–603 (F1)	α1 604–618 (F2)
Brown bear	*in silico*	1593	2792, 2793, 2794	1590 or 1606	1281
Cave bear	E-3-89	1592	2793	x	1281
Cave bear	E-VS-55	1592	2794	1606	x
Cave bear	LIN-E-234	1592	2794	1606	x
Cave bear	VL-L-549	1592	2793	1606	x
Cave bear	LIN-1009	1592	2793	1606	x
Cave bear	AX1069	1592	2793	x	1281
Cave bear	AX4119	1592	2793	1606	x
Cave bear	ERM404	1592	2794	1606	1281
Cave bear	ERM405	1592	2794	1606	1281
Cave bear	CHIOA2	1592	2793	1606	x
Cave bear	CO2	1592	2793	x	x
Cave bear	LL2	1593	2793	1606	x
Cave bear	LL1	1592	2793	x	x
Cave bear	SW3	1592	2794	1606	x
Cave bear	PZ2	1592	2793	x	x
Cave bear	KIZ1	1592	2793	x	x
Cave bear	KIZ4	1592	2793	x	x
Cave bear	MED1	x	2793	x	x
Cave bear	MED2	1592	2793	x	x
Cave bear	MED3	1592	2793	x	x
Brown bear	ARLU39	1592	2793	1606	x
Brown bear	ARLU42	1592	2793	1606	x
Brown bear	SIPA61	1592	2793	x	1281
Brown bear	SIPA215	1592	2793	x	x
Brown bear	PURLU9	1592	2793	1590	x
Brown bear	RT01	x	x	x	x
Brown bear	CGLL51	1592	2793	x	x
Brown bear	S1SH5	1593	2793	x	x
Brown bear	GP-1	x	x	x	x
Brown bear	BEL1	1592	x	1606	x

both prolines hydroxylated. In addition, the existence of a K residue before a G (before tryptic digestion) may add an extra hydroxylation, which would yield an *m*/*z* value of 1606.8.

The second peptide, or F2 for short, has a variable sequence, depending on the taxa. In the review of *in silico* sequences we have found up to three variants, produced by the substitution of an amino acid in the third position of the peptide. In some taxa such as *Ovis* and *Equus*, the third amino acid is T and the mass of the peptide is 1311.6. In *Homo sapiens* the third place is occupied by S and the mass is 1297, as also occurs in other primates as well as in rhinoceros, hippopotamus and some seals. In most carnivores including bears, canids, felids and mustelids, the sequence is GEAGPSGPAGPTGAR and the *m*/*z* value is 1281.6. It does not present any amino acids susceptible to post-translational modifications. This same sequence is found also in *Bos*.

The sum of *m*/*z* of both peptides does not correspond to the value reported in the databases. This is due to the fact that the peptide bond established between the carboxyl group (–COOH) of an amino acid and the amino group (–NH2) of the immediately adjacent amino acid leads to the release of a water molecule (−18 Da). Therefore, the *m*/*z* value attributed in previous literature has a mass of 18 Da less than the sum of its two components.

According to this, in the peptide spectrum of a sample, these two peptides should be either found separately, or joined together if the action of trypsin failed to separate them. Whether or not the amino acid sequence is cleaved between residues 603 and 604 may be due to the performance of the trypsin used, or even to the digestion time. As the trypsin used for all our samples is the same, it can be estimated that it is the digestion time that determines the performance of tryptic cleavage of this fraction of the collagen molecule. In 11 of the cave bear samples studied in this work a peptide 1606 is detected, but only in four samples a peak 1281 appears. In four other samples there is a peptide 2853. In the brown bear set, the peak corresponding to the two peptides together appears in only two samples, while peak 1606 is in three (and in one more, the 1590 variant). Peak 1281, which would correspond to the second peptide fraction, appears in only one case. The scarce occurrence of this peak allows us to hypothesise that the theoretical sequence obtained from the *in-silico* analysis may not be accurate, which is not surprising since there is only one sequence of brown bear col1α1, recalling it was not curated.

Finally, Table 5 shows the presence of peaks in the analysed samples at the *m*/*z* values calculated for these three peptides from the *in-silico* analysis of the UniProtKB brown bear sequence. None of them is crucial for the identification of bear skeletal remains by ZooMS if the collagen is well preserved. However, the E marker (α2 454–483) may be useful for differentiating between ursids and felids when the larger peptide G (COLα2 767–799) is not present.

In any case, the final identification of the peptides must be carried out by other proteomic techniques, such as liquid chromatography–mass spectrometry. It is necessary to ensure that the peaks found in the spectra of the analysed bears really correspond to the sequences in the established position, and that they are not the result of *m*/*z* coincidence with other peptides or peptide fragments present in the collagen.

4.2. Comparison of sequences between brown bear and cave bear species

We found no differences between the peptide markers of brown bear and cave bear of any species, at least not for the commonly used peptide markers. This is not surprising, as the divergence

https://doi.org/10.1017/S1755691023000038 Published online by Cambridge University Press

between brown bears and all cave species has been set at 1.5 million years (based on nuclear DNA, Barlow *et al*. 2021). As we have already seen, the need to maintain the stability of the collagen molecule prevents major changes in its amino acid sequence. The rate of amino acid substitution is estimated to be one every 1–8 million years depending on the vertebrate class (Buckley 2018). Therefore, taxonomic identification based on ZooMS generally does not achieve more than genus rank. Also, genetic divergence between cave bear species occurred less than one million years ago (Barlow *et al*. 2021), which would not be sufficient to cause amino acid substitutions in collagen molecules, at least in the peptide markers used. For example, the oldest separation within the cave bear lineage was about 0.83 million years ago between *U. rossicus* (represented in our samples by those from Kizel in the Urals) and all others (Barlow *et al*. 2021) with no differences in the peptide markers commonly used. Nevertheless, there is a possibility that there are substitutions in other regions of the collagen molecules, but to detect these, the entire collagen molecules would have to be sequenced, as single substitutions cannot be detected simply by their *m*/*z* values.

4.3. Sequence variability between ursids of different chronology

The samples studied range in age from a present-day brown bear, several Holocene brown bears to one brown bear more than 40,000 years BP, and several Pleistocene cave bears of different ages. It might be expected that the collagen from the older bears would be more degraded and have fewer marker peptides due to fragmentation of the molecule during diagenesis, but this is not a particularly visible effect (see Tables 4 and 5). Similarly, there is no pattern in the presence or modification of peptides according to their geographical origin. We might also expect more instances of Q deamidation in the older samples, as this is a known diagenetic process that was proposed as an indicator of the age of skeletal remains (Wilson *et al*. 2012). In peptide G or COL1α2 767–799 this type of degradation is visible in a +1 Da mass shift that seems to affect cave bear samples more than brown bear samples, although with little difference. Certainly, the extant brown bear sample did not show this increase in the mass of peak 2957. But in Eirós, whose two cave bears are separated by 20,000 years, the sample with deamidation is the most modern.

This is consistent with the observation that the extent of deamidation seems to be influenced more by burial conditions than chronological age (Van Doorn *et al*. 2012; Schroeter & Cleland 2015; Welker *et al*. 2017). We did not really expect much degradation in the extracted collagen, as all fossil bones come from cave deposits. Caves maintain fairly stable conditions of humidity and temperature, so the preservation of organic molecules can be sustained over time (Pinto-Llona *et al*. 2005; Torres *et al*. 2014; González-Fortes, *et al*. 2017).

5. Conclusions

The sequence of amino acids that make up the bone collagen molecule shows slight variations between taxa, that can be studied by peptide mass fingerprinting for taxonomic purposes. This requires reference databases that allow peptide identification. In this work we specifically review the identification of peptide markers in cave bears and brown bears, common components of the European Pleistocene fauna. We provide the peptide spectra of 20 cave bear and 10 brown bear samples, that have never been published before. We found no differences between the different cave bear species and brown bear spectra, at least not for the peptide markers described in the literature. We also found no evident correlation between the age of the samples and the post-depositional alterations of their collagen.

However, the *in-silico* study of the ursid collagen sequences published in UniProtKB revealed discrepancies in *m*/*z* values of some peptides, keeping in mind that the sequences are automatically generated and must be annotated. This means that they may contain errors. In particular, the peptides COLα2 292–309 (P2) and COLα2 454–483 (E) have different *m*/*z values* than those suggested for ursids in publications on the subject. These markers are also not found in many published fossil Ursidae spectra. Our analysis of the peptide spectra of 30 samples morphologically identified as ursids revealed a higher affinity for these markers to the values obtained from the *in-silico* analysis (in 27 of the 30 specimens studied, for both P2 and E peptide markers), suggesting that these two peptide markers are not well defined for ursids.

A third tryptic peptide raises problems, that are methodological in nature in this case. The peptide COL1α1 586–618 (F) is actually composed of two tryptic peptides that may or may not be cleaved, apparently depending either on the type of trypsin used or the digestion time, among other possible factors. In the samples analysed here, this peptide is barely detected; however, one of the peptides that we identified as one of its components appears in more than half of the samples. These types of limitations are common in a technique as young as ZooMS. It would be necessary to establish a standardised analysis protocol to avoid such discrepancies.

6. Supplementary material

Supplementary material is available online at https://doi.org/10.1017/S1755691023000038.

7. Acknowledgements

The authors are grateful to Dr G. Rabeder, Dr I. Martini and Dr G. Baryshnikov for providing cave bear samples from Austria, Italy and Russia for analysis, and to Dr JF. García Marín for the modern brown bear sample. ACPLL, JMG, TT, AGV and AGD obtained the samples from the excavation of the sites and made the morphological identification of the skeletal remains and its initial dating. AGV and AGD performed the collagen analysis and interpretation of the results. AGV and AGD wrote the text, with the participation of all authors in the final revision. The authors are grateful for the valuable comments of two anonymous reviewers. Funding for open access charge: Universidade da Coruña/CISUG.

8. Financial support

This study was carried out with the financial support of the project ED431B 2021/17 of the Autonomous Government of Galicia (Spain) awarded to AGD.

9. Competing interest

None.

10. References

Ambrose, S. H. 1990. Preparation and characterization of bone and tooth collagen for isotopic analysis. *Journal of Archaeological Science* **17**, 431–51.

Barlow, A., Paijmans, J. L. A., Alberti, F., Gasparyan, B., Bar-Oz, G., Pinhasi, R., Foronova, I., Puzachenko, A. Y., Pacher, M., Dalén, L., Baryshnikov, G. & Hofreiter, M. 2021. Middle Pleistocene genome calibrates a revised evolutionary history of extinct cave bears. *Current Biology* **31**, 1771–9.

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). *Quaternary Research* **48**, 370–80.

Britton, K., Gaudzinski-Windheuser, S., Roebroeks, W., Kindler, L. & Richards, M. P. 2012. Stable isotope analysis of well-preserved 120,000-year-old herbivore bone collagen from the Middle Palaeolithic site of Neumark-Nord 2, Germany reveals niche separation between bovids and equids. *Palaeogeography, Palaeoclimatology, Palaeoecology* **333**, 168–77.

Brown, S., Douka, K., Collins, M. & Richter, K. K. 2021a. On the standardization of ZooMS nomenclature. *Journal of Proteomics* **235**, 104041.

Brown, S., Wang, N., Oertle, A., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Comeskey, D., Jope-Street, B., Harvey, V. L., Chowdhury, M. P., Buckley, M., Higham, T. & Douka, K. 2021b. Zooarchaeology through the lens of collagen fingerprinting at Denisova Cave. *Scientific Reports* **11**, 15457.

Buckley, M. 2018. Zooarchaeology by mass spectrometry (ZooMS) collagen fingerprinting for the species identification of archaeological bone fragments. In Giovas, C. M. & LeFebvre, M. J (eds) *Zooarchaeology in Practice*, 227–47. Springer.

Buckley, M. & Cheylan, M. 2020. Collagen fingerprinting for the species identification of archaeological amphibian remains. *Boreas* **49**, 709–17.

Buckley, M. & Collins, M. J. 2011. Collagen survival and its use for species identification in Holocene–Lower Pleistocene bone fragments from British archaeological and paleontological sites. *Antiqua* **1**, 1–7.

Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **23**, 3843–54.

Buckley, M., Cosgrove, R., Garvey, J. & Prideaux, G. J. 2017a. Identifying remains of extinct kangaroos in Late Pleistocene deposits using collagen fingerprinting. *Journal of Quaternary Science* **32**, 653–60.

Buckley, M., Fraser, S., Herman, J., Melton, N. D., Mulville, J. & Pálsdóttir, A. H. 2014. Species identification of archaeological marine mammals using collagen fingerprinting. *Journal of Archaeological Science* **41**, 631–41.

Buckley, M., Gu, M., Shameer, S., Patel, S. & Chamberlain, A. T. 2016. High-throughput collagen fingerprinting of intact microfaunal remains; a low-cost method for distinguishing between murine rodent bones. *Rapid Communications in Mass Spectrometry* **30**, 805–12.

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**, 402–11.

Buckley, M. & Kansa, S. W. 2011. Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. *Archaeological and Anthropological Sciences* **3**, 271–80.

Crockett, J. C., Rogers, M. J., Coxon, F. P., Hocking, L. J. & Helfrich, M. H. 2011. Bone remodelling at a glance. *Journal of Cell Science* **124**, 991–8.

Davies, J. E. & Hosseini, M. M. 2000. Histodinamics of endosseous would healing. In Davies, J. E. (ed.) *Bone Engineering*, 1–14. Toronto: EM Squared Incorporated.

De Lombera, A. 2020. *Gestión de materiales silíceos macrocristalinos en el Pleistoceno de Europa occidental. Evolución de las estrategias de adaptación a los condicionantes litológicos* [Management of macrocrystalline siliceous materials in the Pleistocene of Western Europe. Evolution of adaptation strategies to lithological conditions.] PhD Thesis, Universidade de Santiago de Compostela, Spain. [In Spanish.]

DeNiro, M. J. 1985. Post-mortem preservation and alteration of *in vivo* bone collagen isotope ratios in relation to paleodietary reconstruction. *Nature* **317**, 806–9.

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? *Archaeological and Anthropological Sciences* **1**, 31–42.

Döppes, D., Rabeder, G., Frischauf, C., Kavcik-Graumann, N., Kromer, B. & Lindauer, S. 2019. Extinction pattern of Alpine cave bears – new data and climatological interpretation. *Historical Biology* **31**, 422–8.

Eastoe, J. E. 1955. The amino acid composition of mammalian collagen and gelatin. *Biochemical Journal* **61**, 589–600.

Fernández Mosquera, D., Vila-Taboada, M. & Grandal-d'Anglade, A. 2001. Stable isotopes ($\delta^{13}C$, $\delta^{15}N$) from cave bear (*Ursus spelaeus*). A new approach to its paleoenvironment and dormancy. *Proceedings of the Royal Society of London B* **68**, 1159–64.

García-Vázquez, A., Grandal-d'Anglade, A., Vaqueiro Rodríguez, M. & Vidal Romaní, J. R. 2011. On the relation between cave and brown bears in O Courel Sierra (Galicia, NW Spain). *Quaternaire, Hors-série* **4**, 59–69.

García-Vázquez, A., Pinto-Llona, A. C. & Grandal-d'Anglade, A. 2018. Brown bear (*Ursus arctos* L.) paleoecology in the Upper Pleistocene and Holocene of the 695 NW of the Iberian Peninsula: a stable isotope study. *Quaternary International* **481**, 42–51.

García-Vázquez, A., Pinto Llona, A. C., González-Fortes, G. M. & Grandal-d'Anglade, A. 2015. Distribution and chronology of brown bear (*Ursus arctos* L.) in the Iberian Peninsula during Upper Pleistocene and Holocene. *Spanish Journal of Paleontology* **30**, 161–84.

González-Fortes, G., García-Vázquez, A., Pinto Llona, A. C. & Grandal-d'Anglade, A. 2017. Estudio de la preservación de ADN en muestras de úrsidos Pleistocenos y Holocenos del NW de la Península Ibérica [Study of the preservation of DNA in samples of Pleistocene and Holocene ursids from the NW of the Iberian Peninsula]. *Cadernos do Laboratorio Xeolóxico de Laxe* **39**, 127–38. [In Spanish.]

González-Fortes, G., Grandal-d'Anglade, A., Kolbe, B., Fernandes, D., Meleg, I., García-Vázquez, A., Pinto-Llona, A. C., Constantin, S., Torres, T. J., Ortiz, J. E., Frischauf, C., Rabeder, G., Hofreiter, M. & Barlow, A. 2016. Ancient DNA reveals differences in behaviour and sociality between brown bears and extinct cave bears. *Molecular Ecology* **25**, 4907–18.

Grandal-d'Anglade, A., Pérez-Rama, M., García-Vázquez, A. & González-Fortes, G. 2019. The cave bear's hibernation: reconstructing the physiology and behaviour of an extinct animal. *Historical Biology* **31**, 429–41.

Harvey, V. L., Daugnora, L. & Buckley, M. 2018. Species identification of ancient Lithuanian fish remains using collagen fingerprinting. *Journal of Archaeological Science* **98**, 102–11.

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 Society Open Science* **6**, p.191137. https://doi.org/10.1098/rsos.191137

Henriksen, K. & Karsdal, M. A. 2019. Chapter 1 – type I collagen. In Karsdal, Morten A. (ed.) *Biochemistry of Collagens, Laminins and Elastin*, 2nd edn, 1–12. London: Academic Press.

Horn, I. R., Kenens, Y., Palmblad, N. M., van der Plas-Duivesteijn, S. J., Langeveld, B. W., Meijer, H. J. M., Dalebout, H., Marissen, R. J., Fischer, A., Vincent Florens, F. B., Niemann, J., Rijsdijk, K. F., Schulp, A. S., Laros, J. F. J. & Gravendeel, B. 2019. Palaeoproteomics of bird bones for taxonomic classification. *Zoological Journal of the Linnean Society* **186**, 650–65.

James, P., Quadroni, M., Carafoli, E. & Gönnet, G. 1993. Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications* **195**, 58–64.

Katzenberg, M. A. & Waters-Rist, A. 2019. Stable isotope analysis: a tool for studying past diet, demography, and life history. In Katzenberg, M. A. y Grauer, A. (eds) *Biological Anthropology of the Human Skeleton*, 467–504. New York: John Wiley & Sons, Inc.

Kirby, D. P., Buckley, M., Promise, E., Trauger, S. A. & Holdcraft, T. R. 2013. Identification of collagen-based materials in cultural heritage. *The Analyst* **138**, 4849–58.

Kuhn, K. 1987. The classical collagens: types I, II, and III. In Mayne, R. and Burgeson, R. (eds) *Structure and Function of Collagen Types*, 1–42. Academic Press.

Lee, Y. C., Chiang, C. C., Huang, P. Y., Chung, C. Y., Huang, T. D., Wang, C. C., Chen, C. I., Chang, R. S., Liao, C. H. & Reisz, R. R. 2017. Evidence of preserved collagen in an Early Jurassic sauropodomorph dinosaur revealed by synchrotron FTIR 732 microspectroscopy. *Nature Communications* **8**, 14220.

Linnaeus, C. 1758. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis [The system of nature through the three kingdoms of nature, according to classes, orders, genera, species, with characters, differences, synonyms, places]. Laurentii Salvii. [In Latin.]

Longin, R. 1971. New method of collagen extraction for radiocarbon dating. *Nature* **230**, 241–2.

Magrane, M. & UniProt Consortium. 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database*, **2011**, bar009.

Maroto Genover, J. 1993. La Cueva de los Ermitons (Sales de Llierca, Girona): un yacimiento del Paleolítico Medio final. Espacio Tiempo y Forma. *Serie I, Prehistoria y Arqueología* **6**, 13–30. https://doi.org/10.5944/etfi.6.1993.4579

Maroto, J., Vaquero, M., Arrizabalaga, A., Baena, J., Baquedano, E., Jordá, J., Julià, R., Van der Plicht, J., Rasines, P. & Woord, R. 2012. Current issues in late Middle Palaeolithic chronology: new assessments from Northern Iberia. *Quaternary International* **247**, 15–25.

Martini, I., Coltorti, M., Mazza, P. P., Rustioni, M. & Sandrelli, F. 2014. The latest *Ursus spelaeus* in Italy, a new contribution to the extinction chronology of the cave bear. *Quaternary Research* **81**, 117–24.

Olsen, J. V., Ong, S. E. & Mann, M. 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & Cellular Proteomics* **3**, 608–14.

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. *Geochimica et Cosmochimica Acta* **64**, 1043–50.

Pacher, M. & Stuart, A. 2009. Extinction chronology and palaeobiology of the cave bear (*Ursus spelaeus*). *Boreas* **38**, 189–206.

Pappin, D. J., Hojrup, P. & Bleasby, A. J. 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* **3**, 327–32.

Pérez-Rama, M., Fernández-Mosquera, D. & Grandal-d'Anglade, A. 2011. Effects of hibernation on the stable isotope signatures of adult and neonate cave bears. *Quaternaire Hors-série* **4**, 79–88.

Peters, C., Richter, K. K., Manne, T., Dortch, J., Paterson, A., Travouillon, K., Louys, J., Price, G. J., Petraglia, M., Crowther, A. & Boivin, N. 2021. Species identification of Australian marsupials using collagen fingerprinting. *Royal Society Open Science* **8**, 211229.

Pinto-Llona, A. C., Andrews, P. J. & Etxeberría, F. 2005. *Tafonomía y Paleoecología de úrsidos Cuaternarios Cantabricos* [Taphonomy and paleoecology of Cantabrian Quaternary ursids]. Fundación Oso de Asturias, Oviedo. [In Spanish.]

Presslee, S., Wilson, J., Woolley, J., Best, J., Russell, D., Radini, A., Fischer, R., Kessler, B., Boano, R., Collins, M. & Demarchi, B. 2017. The identification of archaeological eggshell using peptide markers. *The Star* **3**, 89–99.

Ricard-Blum, S. 2011. The collagen family. *Cold Spring Harbor Perspectives in Biology* **3**, a004978.

Richter, K. K., Codlin, M. C., Seabrook, M. & Warinner, C. 2022. A primer for ZooMS applications in archaeology. *Proceedings of the National Academy of Sciences of the United States of America* **119**, e2109323119.

Richter, K. K., Wilson, J., Jones, A. K. G., Buckley, M., van Doorn, N. & Collins, M. J. 2011. Fish'n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments. *Journal of Archaeological Science* **38**, 1502–10.

Robinson, A. B. & Rudd, C. J. 1974. Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. *Current Topics in Cellular Regulation* **8**, 247–95.

Rosenmüller, J. C. 1794. Quaedam de ossibus fossilibus animalis cuiusdam, historiam eius et cognitionem accuratiorem illustrantia, dissertatio, quam d. 22. Octob. 1794 ad disputandum proposuit Ioannes Christ [Some of the fossil bones of a certain animal, illustrating its history and more accurate knowledge, a dissertation, which d. 22 October 1794 John Christ proposed for discussion]. Rosenmüller Heßberga-Francus, LL.AA.M. In: Theatro anatomico Lipsiensi Prosector assumto socio Io. Chr. Aug. Heinroth Lips. Med. Stud. Cum tabula aenea. 34 S, pp. 1–34. [In Latin.]

Schroeter, E. R. & Cleland, T. P. 2015. Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Communications in Mass Spectrometry* **30**, 251–525.

Schweitzer, M. H., Zheng, W., Organ, C. L., Avci, R., Suo, Z., Freimark, L. M., Lebleu, V. S., Duncan, M. B., Vander Heiden, M. G. & Neveu, J. M. 2009. Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. canadensis*. *Science* **324**, 626–31.

Shoulders, M. D. & Raines, R. T. 2009. Collagen structure and stability. *Annual Reviews in Biochemistry* **78**, 929–58.

Stewart, J. R. M., Allen, R. B., Jones, A. K. G., Penkman, K. E. H. & Collins, M. J. 2013. ZooMS: making eggshell visible in the archaeological record. *Journal of Archaeological Science* **40**, 1797–804.

Torres, T., Ortiz, J. E., Cobo, R., De Hoz, P., García-Redondo, A. & Grün, R. 2007. Hominid exploitation of the environment and cave bear populations. The case of *Ursus spelaeus* Rosenmüller-Heinroth in Amutxate cave (Aralar, Navarra, Spain). *Journal of Human Evolution* **52**, 1–15.

Torres, T., Ortiz, J. E., Fernández, E., Arroyo-Pardo, E., Grün, R. & Pérez-González, A. 2014. Aspartic acid racemization as a dating tool for dentine: a reality. *Quaternary Geochronology* **22**, 43–56.

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 Communications in Mass Spectrometry* **26**, 2319–27.

Van Klinken, G. J. 1999. Bone collagen quality indicators for palaeodietary and radio-carbon measurements. *Journal of Archaeological Science* **26**, 687–95.

Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B. M., Stewart, J. R., Pääbo, S., Collins, M. J. & Jean-Jacques Hublin, J.-J. 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 11162–7.

Welker, F., Soressi, M., Roussel, M., van Riemsdijk, I., Hublin, J.-J. & Collins, M. J. 2017. Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. *Science and Technology in Archaeological Research (STAR)* **3**, 15–27.

Wilson, J., van Doorn, N. L. & Collins, M. J. 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen. *Analitical Chemistry* **84**, 9041–8. 807.

Yamauchi, M. & Sricholpech, M. 2012. Lysine post-translational modifications of collagen. *Essays in Biochemistry* **52**, 113–33.

MS received 30 May 2022. Accepted for publication 25 January 2023. First published online 27 February 2023