



OPEN Palaeoproteomic insights into the evolutionary history of cave bears from southwestern Europe

Amanda Gutiérrez-Carbajal^{1,3,✉}, Elena Santos^{2,3,4}, Esther Lizano^{1,5,6}, Tomas Marques-Bonet^{1,5,7,8}, Ricardo Fong-Zazueta⁹, Aurora Grandal-d'Anglade¹⁰, David M. Alba^{5,6}, Ana García-Vázquez¹¹, Asier Gómez-Olivencia^{2,12,13}, Mónica Villalba de Alvarado^{2,13,14}, Luca Pandolfi¹⁵, Lorenzo Rook¹⁶, Montserrat Sanz^{17,18}, Joan Daura^{17,18}, João Zilhão¹⁸, Juan Luis Arsuaga^{2,19}, José María Bermúdez de Castro³ & María Martínón-Torres^{3,20}

Dental enamel proteomics provides access to molecular information from time periods far beyond the reach of ancient DNA, offering a powerful means to investigate evolutionary dynamics in deep time. Here we present the first large-scale enamel proteomic dataset for 55 cave bears to date, analysing fossils spanning the Early to Late Pleistocene across southwestern Europe, including key specimens from Atapuerca sites. Using a non-enzymatic demineralization workflow and LC-MS/MS analysis we consistently recovered rich enamel proteomes and identified taxonomically informative peptides alongside two previously unknown amino acid variations in AMBN-249 and SERPINA1-341 positions. Phylogenetic analyses are consistent with a basal position of *Ursus dolinensis* within the speleoid clade, providing new molecular evidence relevant to its debated taxonomy. Together, these findings show that enamel proteomes preserve phylogenetic structure and lineage-specific molecular signals, highlighting the potential of palaeoproteomics to inform evolutionary relationships within the speleoid clade across the Pleistocene.

Keywords Proteome, *Ursus spelaeus*, Pleistocene, Enamel, Proteins

The evolutionary history of cave bears (the speleoid lineage) has long been a subject of extensive paleontological and genetic interest due to their extensive fossil record and ecological significance during the Pleistocene. These ursids, distributed from the Iberian Peninsula across Eurasia to north eastern Siberia¹, are predominantly known

¹Departament de Medicina i Ciències de la Vida, Institut de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain. ²Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, Madrid, Spain. ³CENIEH (Centro Nacional de Investigación sobre la Evolución Humana), Burgos, Spain. ⁴Departamento de Ciencias de la Vida, Cátedra de Otoacústica Evolutiva y Paleoantropología (HM Hospitales-Universidad de Alcalá), Universidad de Alcalá, Alcalá de Henares, Madrid, Spain. ⁵Institut Català de Paleontologia Miquel Crusafont (ICP-CERCA), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain. ⁶Unidad de Paleobiología, ICP-CERCA, Unidad Asociada al CSIC por el IBE UPF-CSIC, Cerdanyola del Vallès, Barcelona, Spain. ⁷Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, Barcelona, Spain. ⁸CNAG, Centro Nacional de Análisis Genómico, Barcelona, Spain. ⁹Département de sciences biologiques, Université de Montréal, Montréal, Canada. ¹⁰Instituto Universitario de Xeoloxía, Universidade da Coruña, ESCI, A Coruña, Spain. ¹¹ArchaeoScience Platform (ASp), Research Institute of the University of Bucharest (ICUB), University of Bucharest, Bucharest, Romania. ¹²Departamento de Geología, Facultad de Ciencia y Tecnología, Universidad del País Vasco-Euskal Herriko Unibertsitatea (UPV/EHU), Leioa, Spain. ¹³Sociedad de Ciencias Aranzadi, Donostia-San Sebastián, Spain. ¹⁴Departamento de Procesos Psicológicos Básicos y su Desarrollo, Universidad de País Vasco-Euskal Herriko Unibertsitatea (UPV/EHU), Leioa, Spain. ¹⁵Dipartimento di Scienze della Terra, Università di Pisa, Pisa, Italy. ¹⁶Dipartimento di Scienze della Terra, Paleo[Fab]Lab, Università degli Studi di Firenze, Firenze, Italy. ¹⁷Departament d'Història i Arqueologia, Grup de Recerca del Quaternari (GRQ)-SERP, Universitat de Barcelona, Barcelona, Spain. ¹⁸Faculdade de Letras, UNIARQ-Centro de Arqueologia da Universidade de Lisboa, Lisboa, Portugal. ¹⁹Departamento de Geodinámica, Estratigrafía y Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, Madrid, Spain. ²⁰Department of Anthropology, University College London, London, UK. ✉email: amanda.gutierrez@upf.edu

from karstic deposits, reflecting their hibernation behaviour^{2–4}, which is also known to leave a characteristic isotopic imprint due to metabolic shifts during dormancy⁵. Phylogenetically, cave bears (*Ursus deningeri*–*Ursus spelaeus sensu lato*) form a sister clade to the extant brown (*Ursus arctos* Linnaeus, 1758) and polar bears (*Ursus maritimus* Phipps, 1774), diverging approximately 1.2–1.6 million years ago (Ma) based on nuclear and mitochondrial DNA clocks^{6–9}. The lineage persisted until ~24–26 thousand years ago (ka), when the last representatives became extinct, likely following a demographic decline possibly influenced by climatic changes and human competition for cave habitats^{10–12}.

Multiple hypotheses have been proposed regarding the origin and diversification of this lineage. Morphological research, from broad biochronological syntheses to detailed population-level analyses, has alternately supported gradual (anagenetic) evolution or more complex branching scenarios^{13–16}. *Ursus etruscus* Cuvier, 1823 is often regarded as the common ancestor of both cave and brown bears^{2,3,17}. Ancient DNA (aDNA) analyses have clarified relationships among Late Pleistocene taxa (*Ursus spelaeus* Rosenmüller, 1794, *Ursus kanivetz* Vereshchagin, 1973, and *U. kudarensis* Baryshnikov, 1985) revealing regional structuring and divergence^{9,18}, but their temporal reach rarely extends beyond the Middle Pleistocene^{9,19} due to preservation limits. The oldest mitochondrial sequences currently available derive from *Ursus deningeri* von Reichenau, 1904 specimens from Sima de los Huesos (~430–300 ka)^{20,21}, a well-constrained Middle Pleistocene site²², placing a practical lower bound on the application of aDNA approaches and leaving the earlier evolution of the speleoid clade poorly resolved. This gap is particularly evident for *Ursus dolinensis* García & Arsuaga, 2001 from TD4 level in Gran Dolina (Sierra de Atapuerca, Spain), whose phylogenetic position remains debated: some authors place it at the base of the speleoid lineage²³, while others highlight affinities with archaic arctoid forms based on its mosaic of traits^{16,24,25}. Clarifying the evolutionary depth and relationships of these early representatives therefore requires biomolecular evidence capable of extending beyond the temporal limits of aDNA.

Palaeoproteomics offers a means to bridge this gap. Proteins, being more chemically stable than DNA, can persist for over a million years in mineralized tissues such as dental enamel^{26,27}. Advances in mass spectrometry and bioinformatics now allow recovery of phylogenetically informative single amino acid polymorphisms (SAPs) from fossil proteomes²⁸. This approach has already illuminated aspects of Pleistocene mammals²⁹ and hominin evolution^{30–32}, yet carnivores—and ursids in particular—remain largely unexplored. Here, we present the first comprehensive enamel proteomic analysis of fossil bears across a broad temporal and geographic range, over nearly 1 Ma, across southwestern Europe (Fig. 1) with key fossils from the Atapuerca complex and additional sites. By analyzing proteomes from *U. dolinensis*, *U. deningeri*, and *U. spelaeus* and integrating them with extant *Ursus* and *Ailuropoda* references, we aim to (1) characterize enamel proteome composition and (2) evaluate its phylogenetic utility for clarifying cave bear evolutionary history.

Ancient peptides were successfully recovered from all the analyzed 55 fossils, with peptide-spectrum matches (PSMs) varying across temporal groups. Recovery was generally lower in Early Pleistocene and open-air specimens, whereas Middle Pleistocene samples—particularly those from Sima de los Huesos—reached counts comparable to Late Pleistocene individuals (Fig. 2A–C). An extant specimen of *U. arctos* used as a control yielded fewer PSMs overall, although relative intensities of the peptides remained within the expected range (*SI.II.1 Proteomic Recovery Intensity*). Across fossils, fragment length distributions were dominated by short peptides (8–10 amino acids), consistent with long-term hydrolysis. Middle Pleistocene samples showed the highest average peptide counts, with a maximum of ~390 peptides at a length of 10 amino acids, whereas Late Pleistocene samples preserved a greater proportion of longer sequences (>15 amino acids). Early Pleistocene specimens exhibited more advanced degradation, with lower yields and shorter fragments. Post-translational modifications, particularly extensive deamidation, confirmed the antiquity of all fossil sequences, with values consistently higher than in the modern control.

Between 10 and 16 proteins were confidently identified per specimen. Enamel-specific proteins—including amelogenin (AMELX/AMELY), enamelin (ENAM), ameloblastin (AMBN), amelotin (AMTN), odontogenic ameloblast-associated protein (ODAM), and matrix metalloproteinase-20 (MMP20)—were consistently recovered, often accompanied by plasma proteins such as albumin (ALB), alpha-2-HS-glycoprotein (AHSG), alpha-1-antitrypsin (SERPINA1), antithrombin III (SERPINC1), and occasionally by collagen types I and XVII (COL1A1, COL1A2, COL17A1). Less frequent identifications included apolipoprotein A1 (APOA1) and zinc finger protein (ZFC3H1). The combined use of two search engines improved detection but yielded largely overlapping results. Amelogenin isoforms enabled sex determination in most individuals: 27 were identified as male (AMELY-specific peptides), 25 as female or likely female (AMELX-only), and three remained inconclusive. Complete recovery metrics and platform-specific comparisons are provided in *Dataset S2- Peptide Recovery and Identification*. Protein alignments reveal high overall sequence conservation across ursids, yet eight proteins—AHSG, AMBN, ALB, COL17A1, ENAM, MMP20, SERPINA1, and ZFC3H1—contained phylogenetically informative single amino acid polymorphisms (SAPs). In total, 48 SAPs were detected. A detailed summary of all identified SAPs, including their occurrence across specimens, phylogenetic support, and confidence assessment, is provided in *Dataset S3- SAP Identification and Phylogenetic Support*. AMBN and ALB had the highest counts (11 each), followed by COL17A1 and ENAM (9 each). Most reflected intergeneric (*Ursus*–*Ailuropoda*) or intra-*Ursus* variation, and only five appeared potentially unique to extinct cave bears (*SI.II.2 SAP Validation*). After applying strict validation criteria such as requiring consistent support from multiple PSMs and/or overlapping peptides, high-quality MS/MS fragmentation, and agreement across search engines, three of these cave-bear specific SAPs were retained as high-confidence variants.

These novel variants, not previously documented in the reference proteomes of ursids, were detected in AMBN and SERPINA1. A substitution of tyrosine (Y) to serine (S) at position AMBN-249 was consistently present in *U. deningeri* and *U. spelaeus* specimens from multiple sites in Spain, Portugal, and Italy (Fig. 3). Two Middle Pleistocene individuals from the level TD8 (Gran Dolina, Spain) and Aroeira (Portugal) presented ambiguous signals in this position, suggesting incomplete fixation of this variant or potential heterozygosity

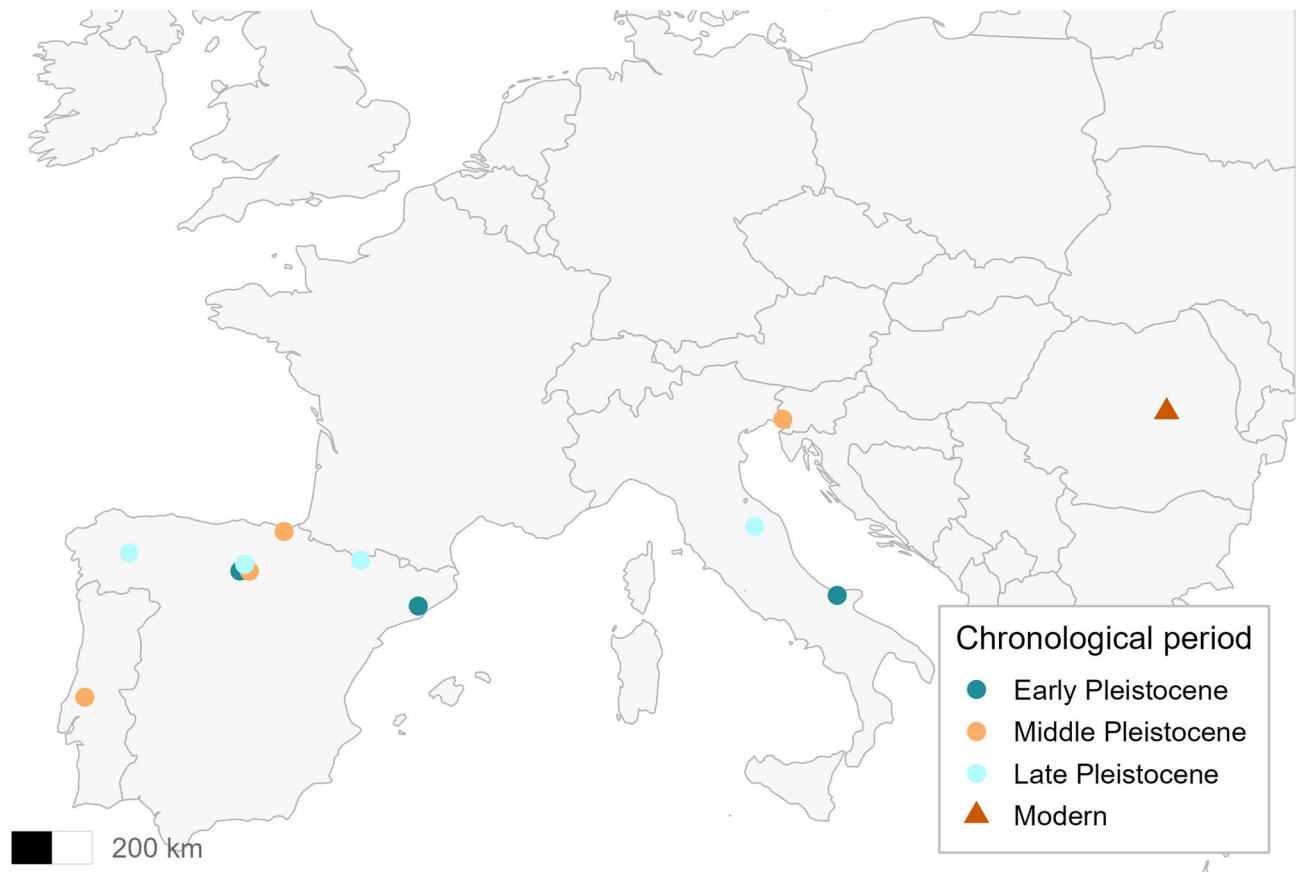


Fig. 1. Spatial and temporal framework of the cave bear specimens analysed in this study. Sampling localities across southwestern Europe, including the Atapuerca sites, and their assignment to Early, Middle and Late Pleistocene periods. The modern *Ursus arctos* specimen is included as a comparative reference. Detailed metadata are available in [Dataset S1- Table S1.1 Sample Metadata](#).

(details in *SI.II.2 SAP Validation*). A second AMBN substitution, leucine (L) to valine (V) at AMBN-278, was observed in a single specimen from Sala Fantasma and is reported cautiously. Both AMBN variants occur in exon 13, a variable region among mammals, and likely represent conservative changes³³. In SERPINA1, an asparagine (N) to threonine (T) substitution at position 341 was consistently recovered in *U. deningeri* and *U. spelaeus* as well as in one *U. arctos* individual from Monte Cucco (Italy), and absent in *U. dolinensis*, providing an additional marker for the speleoid lineage.

Other non-unique SAPs also proved informative. At ALB-580, *U. dolinensis* and *U. deningeri* shared the ancestral state with several extant ursids, while *U. spelaeus* and *U. arctos* carried the derived variant. At ENAM-269, isoleucine (I) occurred in *U. deningeri*, *U. spelaeus*, and *U. maritimus*, contrasting with asparagine (N) in other bears. In COL17A1-1325, *U. deningeri* and *U. spelaeus* shared methionine (M), whereas *U. arctos* and *U. maritimus* carried threonine (T). Coverage for these sites was incomplete in *U. dolinensis*, preventing a complete assessment. Collectively, these patterns reveal conserved molecular cohesion within the *U. deningeri*–*U. spelaeus* clade, retention of ancestral states in *U. dolinensis*, and several lineage-specific markers with potential for biochronological and phylogenetic applications.

Phylogenetic analyses of enamel proteomes yielded congruent topologies across Bayesian and maximum likelihood methods (Fig. 4). All reconstructions recovered *Ursus* as monophyletic, with a basal divergence between arctoid bears and the speleoid lineage. Within speleoids, *U. dolinensis* was consistently recovered as the earliest diverging member, preceding *U. deningeri* and *U. spelaeus*. Although statistical support for this placement was modest, the topology remained stable across datasets. *Ursus deningeri* and *U. spelaeus* formed a strongly supported clade, confirming their close evolutionary relationship and their more derived status relative to *U. dolinensis*. Bayesian inference produced the best resolved tree, with consistent placement of both extinct and extant taxa. An unexpected result was the placement of *U. maritimus* as sister to the *U. deningeri*–*U. spelaeus* clade instead of clustering with *U. arctos*, as reported in previous studies. This position may reflect either database limitations or convergent molecular signals. Additional analyses and full statistical details are provided in the *SI.II.3 Expanded Phylogenetic Trees*.

Beyond phylogenetic structure, enamel proteomes also reveal patterns of protein preservation and functional composition. Across all 55 specimens, enamel-specific proteins—AMELX, AMBN, AMTN, and ENAM—were consistently identified, confirming their stability and central role in enamel formation, consistent with reports

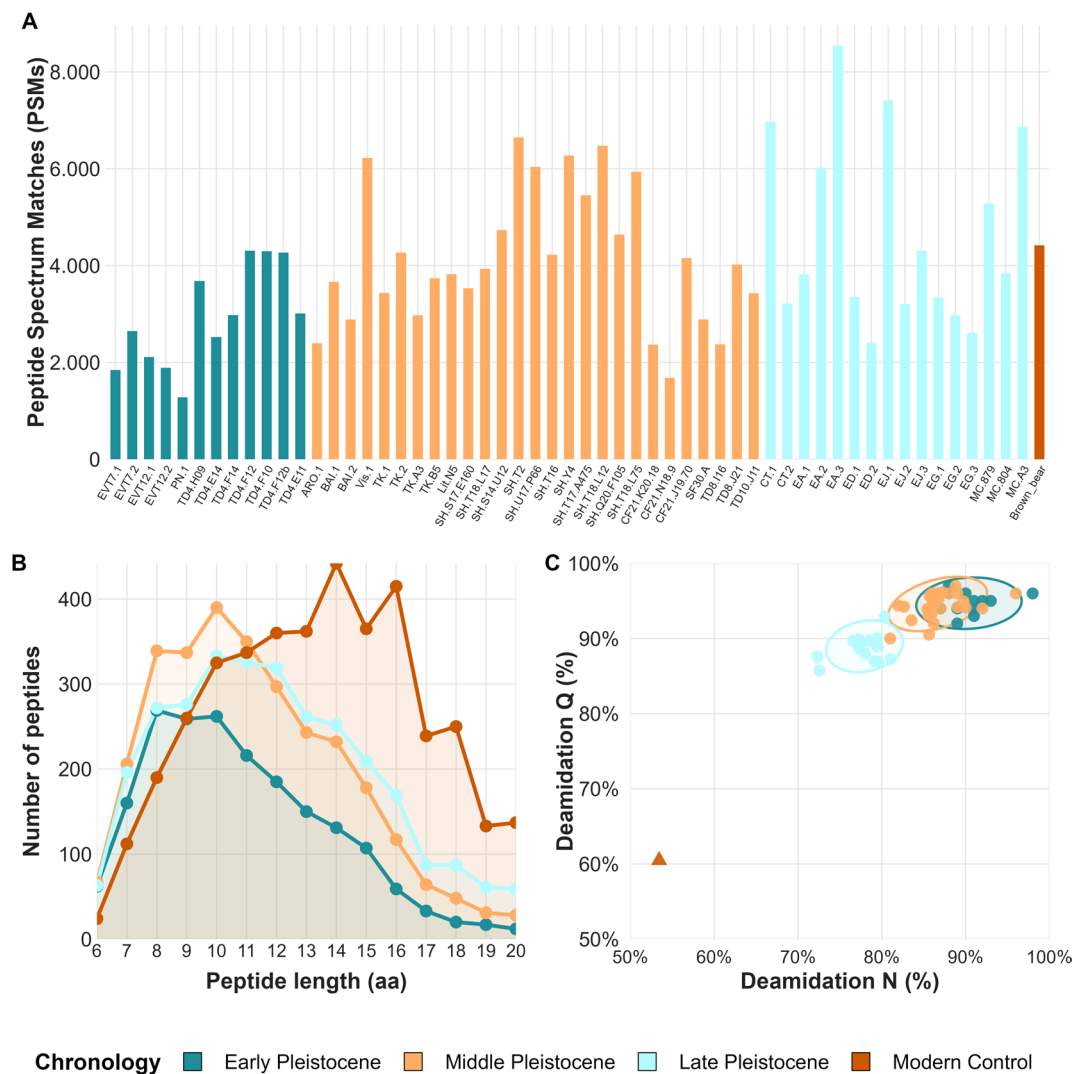


Fig. 2. Endogenous origin and preservation of enamel peptides across temporal groups. **(A)** Proteomic recovery measured as peptide-spectrum matches (PSMs) per specimen. **(B)** Distribution of peptide lengths. Peptide length distributions were calculated by grouping all samples within each chronological period. For each peptide length, peptide counts were averaged within each period. **(C)** Mean rates of deamidation at asparagine/glutamine (N/Q).

from other Pleistocene studies²⁹. Plasma proteins, such as ALB and AHSG, were also commonly recovered; the latter appeared in 43 specimens, supporting its proposed endogenous incorporation into enamel³¹. By contrast, TUFT1 and KLK4, well-known components of mammalian enamel³⁴, were largely absent, suggesting either taxon-specific differences or selective degradation patterns. MMP20 was consistently recovered and exhibited sequence variation, reinforcing its relevance in enamel proteome studies. Collagen peptides were abundant in Late Pleistocene *U. spelaeus* but scarce in older specimens, and no informative variants were observed in COL1 genes, consistent with their strong evolutionary conservation³⁵. ZooMS analyses have similarly reported no COL1 markers distinguishing *U. spelaeus* from *U. arctos*³⁶. In contrast, COL17A1 was detected in nearly all specimens and showed moderate, non-lineage-specific variation. Its essential role in enamel formation, demonstrated by defects in COL17A1-deficient mice³⁷, underscores both its preservation potential and comparative relevance across taxa.

Particularly noteworthy is the recurrent recovery of three serine protease inhibitors (serpins)—alpha-1-antitrypsin (SERPINA1), antithrombin III (SERPINC1), and, more sporadically, the rarely detected pigment epithelium-derived factor (SERPINF1). Specifically, SERPINA1 was detected in nearly all specimens, often with high peptide counts and recurrent post-translational modification patterns, primarily deamidation affecting the same residues across specimens, alongside a lineage-specific SAP. Their consistent presence in enamel proteomes raises questions about the pathways by which they became incorporated. One possibility is a role in amelogenesis, given their co-occurrence with proteases such as MMP20 and KLK4 and previous evidence for interactions with mineralized tissues^{38,39}. Alternatively, their presence may reflect aspects of systemic physiology. In extant animals, including hibernating species, circulating serpins are involved in processes such as metabolic

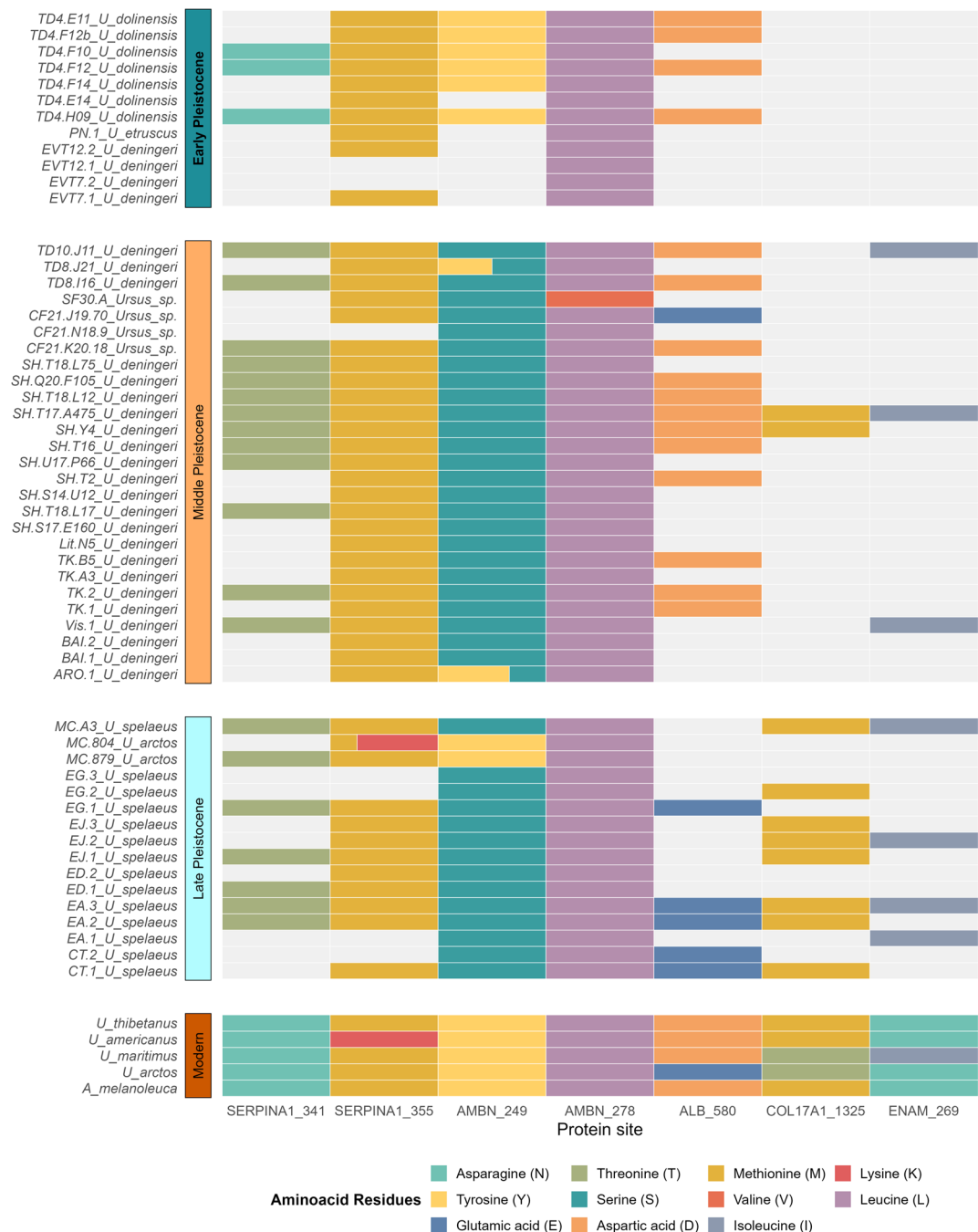


Fig. 3. Sequence variation across enamel and serum proteins in fossil and modern bears. Color-coded matrix displaying informative amino acid positions. Columns represent variable sites and rows correspond to individual specimens. Partial or split-coloured bars indicate heterozygous positions, where two different amino acids were detected in the same protein from a given specimen. The relative proportion of each colour reflects the proportion of peptide-spectrum matches (PSMs) supporting each amino acid variant at that position.

regulation, cardiovascular function and immune modulation^{40–46}, making their repeated detection in cave bear enamel proteomes particularly intriguing.

However, no direct functional inference can be drawn from the present data, and any link to specific physiological states, including hibernation, remains speculative. If enamel mineralization overlapped with periods of physiological change, transiently circulating proteins could potentially have been incorporated into the developing matrix, although this hypothesis requires further investigation.

These observations suggest that enamel proteomes may capture not only phylogenetic signals but also systemically expressed proteins, opening new venues for exploring the biological context of fossil biomolecules. Expanding comparative analyses across both modern and fossil carnivores will be critical to assess whether

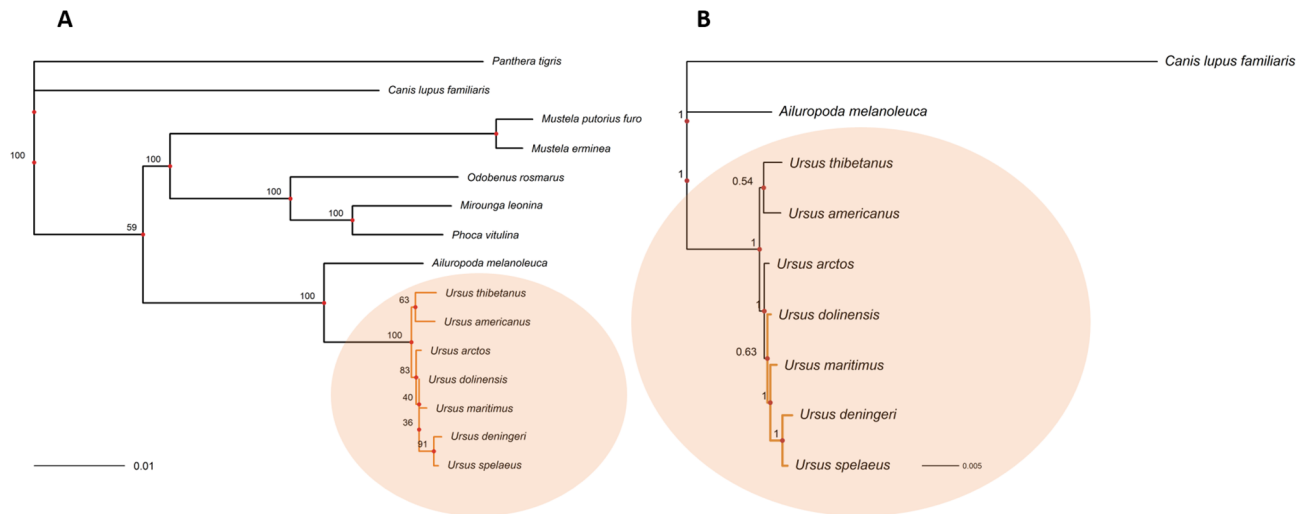


Fig. 4. **A.** Maximum likelihood phylogenetic tree of selected Carnivora, including cave bears, inferred from concatenated protein sequences, enamel-associated and systemic proteins, using IQ-TREE (Order Carnivora dataset). Node values indicate ultrafast bootstrap support (%). **B.** Bayesian phylogenetic tree of Ursidae based on concatenated protein sequences, enamel-associated and systemic proteins, inferred with MrBayes under a mixed amino-acid model across 3,000,000 generations (Family Ursidae dataset). Node values represent posterior probabilities.

these serpins represent lineage-specific traits or more general physiological patterns. Although serpins have been sporadically reported in dental proteomes of non-ursid mammals^{34,47,48}, the frequent detection of SERPINA1 in cave bears points to a potentially distinctive feature of the genus *Ursus*, pending further validation.

Our palaeoproteomic analyses refine phylogenetic relationships within the cave bear lineage. The placement of *U. dolinensis* as basal relative to the *U. deningeri*–*U. spelaeus* clade is consistent with its interpretation as an early member of the speleoid clade, in agreement with some morphology-based phylogenetic hypotheses²³. However, given the limited number of informative positions in enamel proteomes and the geographically restricted nature of the dataset, these results should be interpreted with caution and cannot be readily extrapolated to all *Ursus dolinensis* populations across Europe. Within speleoids, *U. deningeri* and *U. spelaeus* form a strongly supported clade, in line with their close morphological and molecular affinities^{20,21}. Their minimal proteomic distinction, restricted to a single SAP (ALB-580), highlights both their close relationship and the limited sequence divergence detectable in enamel proteins. The possibility that *U. deningeri* represents a direct ancestor of *U. spelaeus* cannot be formally evaluated within the dichotomic phylogenetic framework applied here, although the molecular pattern recovered in this study is consistent with this scenario. Comparisons with extant taxa broadly align with genomic studies: *U. arctos* emerges as the closest living relative of cave bears^{9,49}, while other ursids cluster as expected. The only exception is *U. maritimus*, which appeared as a sister to the *U. deningeri*–*U. spelaeus* clade rather than clustering with *U. arctos*. This placement received weak support and is most plausibly explained by limited peptide coverage or by convergent signals, including the shared state at ENAM-269 (I) observed in both *U. maritimus* and several Middle and Late Pleistocene specimens (see *SI.II.2 SAP Validation*). The lack of peptide coverage at this site in Early Pleistocene individuals, however, limits further evaluation. Together, these observations highlight both the power and the intrinsic constraints of enamel proteomes for phylogenetic inference: while major clades are consistently resolved, the strong conservation of enamel proteins and the fragmentary nature of ancient peptides mean that species-level distinctions can remain challenging⁵⁰. In this context, isolated shared states can occasionally generate locally misleading affinities, but do not alter the overall topology, which remains stable across datasets and analytical frameworks.

Beyond clarifying phylogenetic placement, enamel proteomes reveal evolutionary signatures across both time and space. Two derived variants—AMBN-249(S) and SERPINA1-341(T)—were consistently shared by *U. deningeri* and *U. spelaeus*, providing robust molecular markers for the speleoid lineage. The occasional absence of these markers likely reflects limited peptide recovery rather than genuine biological variation. Their uniform presence across geographically distant sites highlights a remarkable molecular homogeneity in *U. deningeri* proteomes, in contrast with the morphological variability often described for this taxon⁵¹.

Some Middle Pleistocene sites, however, capture intermediate stages. At Aroeira and at level TD8 of the Gran Dolina site, two individuals show both ancestral and derived AMBN-249 states, suggesting partial fixation or population-level heterogeneity. Alternatively, this pattern is compatible with limited gene flow between arctoid and speleoid populations during the early stages of their divergence¹⁸. These observations may reflect a potentially transitional signal. The stratigraphic position of TD8, constrained to postdate the Brunhes–Matuyama reversal identified in the underlying TD7 on the basis of paleomagnetic data⁵², and separated from it by an erosive discontinuity, provides relevant contextual support for its biochronological significance in the context of faunal turnover and ecological restructuring^{53–55}. The Aroeira specimen, although inconclusive, offers a useful contrast with other Middle Pleistocene sites and highlights regional variability in the pace of molecular

change. While the *U. deningeri* proteome appears generally homogeneous, some populations retain ancestral residues, whereas others show early fixation of derived variants. This pattern suggests that molecular transitions within the speleoid lineage were not synchronous across populations, with different groups likely fixing derived variants at different times, potentially shaped by ecological or demographic factors. Such diachronic patterns are consistent with a gradual evolutionary model, although the limited number of specimens and the regional scope of the dataset preclude broader generalisation. However, given that these observations are based on a single amino acid and considering the potential for sampling bias across sites and time intervals, these interpretations should be treated with caution and will require additional specimens to be tested in future research. Additional signals of variability were detected in isolated contexts. At Sala Fantasma, one specimen carried a unique AMBN-278 substitution not seen elsewhere, while others from nearby Cueva Fantasma displayed the typical speleoid signature. At Monte Cucco, two individuals identified morphologically as *U. arctos* yielded proteomic profiles that did not correspond to our *U. arctos* reference, suggesting either misidentification or unresolved variability that can't be confidently interpreted with the current dataset. Finally, the specimen from Pirro Nord (Italy), attributed to *U. etruscus*, was the most degraded in the dataset. Nevertheless, the recovered segments were highly conserved, indistinguishable from both extant and extinct bear species.

Together, these results illustrate the potential of enamel palaeoproteomics to integrate molecular, ecological and evolutionary scales, refining our view of lineage dynamics across the Pleistocene. The persistence of stable molecular signatures, combined with evidence for gradual transitions in key horizons, is consistent with a gradual evolutionary model within the cave bear lineage, although this interpretation remains tentative and should be further tested with expanded datasets.

Methods

Sample selection

We analysed 55 fossil enamel specimens from several sites in southwestern Europe. The dataset includes individuals attributed to *U. dolinensis*, *U. deningeri*, *U. spelaeus*, *U. etruscus*, and *U. arctos*. Molars were targeted because ursids exhibit well-characterised enamel thickness patterns, with cave bears showing markedly thick enamel and particularly robust anterior molars, making these teeth the most suitable elements for enamel proteomic sampling^{56–58}. Particular emphasis was placed on specimens from the Sierra de Atapuerca (Burgos, Spain)—notably Gran Dolina, Sima de los Huesos, and Fantasma sites—which provide key taxonomic and temporal contexts across successive Pleistocene stages. Additional material derives from sites in the Iberian Peninsula (Spain and Portugal) and Italy, representing distinct biochronological intervals of the Pleistocene. A modern *Ursus arctos* individual from Romania was included as a comparative control. Full metadata is provided in *Dataset S1 – Table S1.1 Sample Metadata*.

Enamel sampling

Whenever possible, naturally fractured enamel fragments were prioritised to minimise damage to the specimens. For intact molars, enamel was carefully isolated using sterile tools, avoiding dentine. All manipulations were conducted under laminar-flow conditions using disposable equipment and full protective gear. Enamel fragments (20–120 mg) were cleaned with 10% sodium hypochlorite, rinsed, dried, and stored in sterile tubes. Micro-CT scans were obtained for Atapuerca specimens prior to sampling at the CENIEH-ICTS facility. One extraction blank was included for every five to seven samples.

Protein extraction

Laboratory work was carried out at the Institut de Biologia Evolutiva (IBE, UPF-CSIC) under stringent decontamination protocols. Samples were mechanically powdered with a granite mortar and transferred to Protein LoBind[®] tubes. Demineralization was performed using 10% trifluoroacetic acid (TFA) at 4 °C with agitation for 24 h, repeated for two consecutive cycles to maximise peptide release. After each cycle, supernatants were collected and pellets resuspended in fresh TFA. Peptide desalting was carried out using in-house StageTips⁵⁹ prepared with C18 Empore[™] discs. Tips were conditioned with methanol, then equilibrated with 5% formic acid (FA). Bound peptides were washed in 5% FA, stored at –20 °C, and eluted using 5% FA in 50% acetonitrile (ACN).

LC–MS/MS analysis

Mass spectrometry was performed at the CRG-UPF Proteomics Unit (Barcelona) using an Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific, San Jose, USA) coupled to an EASY-nLC 1200 system. Peptides were separated on a 50 cm × 75 µm C18 column using a linear gradient from 5 to 25% buffer B (0.1% FA in 80% acetonitrile) over 105 min, followed by 25–40% over 15 min and a final wash at 100% buffer B. Flow rate was kept at 300 nL/min. The mass spectrometer operated in positive ion mode (spray voltage 2.4 kV; source temperature 305 °C). Full MS scans were acquired at 120,000 resolution (*m/z* 350–1400) with an AGC target of 4 × 10⁶. Precursors were selected using a Top Speed data-dependent method and fragmented by HCD (NCE = 28%). MS/MS scans were acquired at 30,000 resolution (isolation window 0.7 *m/z*) with dynamic exclusion set to 60 s. Four injection blanks were injected before and after each sample, and digested bovine serum albumin (BSA; NEB cat. #P8108S) was used as a quality control. Instrument performance was monitored with QCloud⁶⁰.

Protein identification

Mass spectrometry data were processed using two complementary proteomic platforms: MaxQuant v2.0.2⁶¹ and PEAKS Studio Xpro (v10.6, Bioinformatics Solutions Inc.). For MaxQuant analysis, raw files were searched against the curated ursid protein databases, each run independently. Searches were performed with non-specific digestion. Variable modifications included deamidation (N/Q), oxidation (M/P), pyro-glutamate formation

(N-term Q to pyro-Glu; N-term E to pyro-Glu) and phosphorylation (S/T/Y). Minimum peptide length was set to 6 amino acids and maximum to 20 aa, with a mass limit of 4,600 Da. The MaxNovo de novo module and dependent-peptide searches were enabled⁶². Proteins and PSMs were filtered at 1% FDR, requiring at least one unique peptide per protein. All remaining settings followed MaxQuant defaults. The same datasets were processed using the full PEAKS workflow (de novo, PEAKS DB and PEAKS SPIDER)^{63,64} to detect peptides containing potential single amino acid polymorphisms (SAPs). Searches were run with no enzyme specificity, 5 ppm precursor tolerance and 0.02 Da fragment tolerance. Variable modifications included deamidation (N/Q), oxidation (M), pyro-glutamate formation (N-term Q to pyro-Glu; N-term E to pyro-Glu), hydroxylation (P) and phosphorylation (S/T/Y), allowing up to three modifications per peptide. Peptide identifications were filtered at 1% FDR, with a minimum protein score of $-10 \lg p \geq 20$ and an ALC% threshold $\geq 95\%$ for de novo spectra.

Custom ursid protein databases were generated for this study and used sequentially during peptide identification. These included curated sets of enamel-associated and systemic proteins from extant *Ursus* and *Ailuropoda*, as well as expanded panels incorporating genome-derived sequences and reconstructed proteins for extinct taxa. All database compositions, reconstruction procedures (including AMELX/AMELY isoforms) and validation steps are fully detailed in *SI. I Additional Methodology* and *Dataset S1 – Tables S1.2–1.4*.

Phylogenetic analysis

Phylogenetic inference was performed using concatenated alignments of 12 enamel-associated and systemic proteins (AHSG, ALB, AMBN, AMELX, AMTN, COL1A1, COL17A1, ENAM, MMP20, ODAM, SERPINA1, SERPINC1). Consensus sequences for each taxon were generated with EMBOSS Cons (European Bioinformatics Institute, EMBL-EBI)⁶⁵, aligned with MAFFT⁶⁶, and filtered with GBlocks⁶⁷ to remove poorly supported regions. Because leucine and isoleucine are isobaric in MS/MS spectra, all variable Isoleucine (I) residues were normalised to Leucine (L). Two comparative datasets were analysed: a broad Carnivora panel and a restricted Ursidae panel, with *Canis lupus familiaris* and *Ailuropoda melanoleuca* as outgroups.

Maximum likelihood trees were inferred in IQ-TREE⁶⁸ using protein-specific substitution models selected by ModelFinder⁶⁹, with node support estimated from 5,000 ultrafast bootstrap replicates. Bayesian inference was conducted in MrBayes⁷⁰ under a mixed amino-acid model across partitions, using four runs of eight chains for 3,000,000 generations. Full methodological details, dataset composition, and model settings are provided in *SI. I Additional Methodology* and *Dataset S1 – Tables S1.5–S1.6*.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE⁷¹ partner repository with the dataset identifier PXD069138. Reconstructed sequences have been deposited in the Zenodo repository under <https://doi.org/10.5281/zenodo.18407395>.

Received: 15 January 2026; Accepted: 11 June 2026

Published online: 24 June 2026

References

- Sher, A. V. et al. The first record of spelaeoid bears in Arctic Siberia. *Q. Sci. Rev.* **30**, 2238–2249 (2011).
- Therrien, E. *Ursidenphylogenie und Biostratigraphie* (Urban und Fischer, 1959).
- Kurtén, B. *Pleistocene mammals of Europe* (Routledge, 1968).
- Kurtén, B. The cave bear story: life and death of a vanished animal. New York. (1976).
- Pérez-Rama, Marta, D. & Fernández-Mosquera Grandal-d'Anglade, A. Effects of hibernation on the stable isotope signatures of adult and neonate cave bears. *Quaternaire* **4**, 79–88 (2011).
- Knapp, M. et al. First DNA sequences from Asian cave bear fossils reveal deep divergences and complex phylogeographic patterns. *Mol. Ecol.* **18**, 1225–1238 (2009).
- Bon, C. et al. Deciphering the complete mitochondrial genome and phylogeny of the extinct cave bear in the Paleolithic painted cave of Chauvet. *Proc. Natl. Acad. Sci.* **105**, 17447–17452 (2008).
- Loreille, O. et al. Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Curr. Biol.* **11**, 200–203 (2001).
- Barlow, A. et al. Middle Pleistocene genome calibrates a revised evolutionary history of extinct cave bears. *Curr. Biol.* **31**, 1771–1779 (2021).
- Baca, M. et al. Retreat and extinction of the Late Pleistocene cave bear (*Ursus spelaeus sensu lato*). *Sci. Nat.* **103**, 92 (2016).
- Pacher, M. & Stuart, A. J. Extinction chronology and palaeobiology of the cave bear (*Ursus spelaeus*). *Boreas* **38**, 189–206 (2009).
- Stiller, M. et al. Withering away—25,000 years of genetic decline preceded cave bear extinction. *Mol. Biol. Evol.* **27**, 975–978 (2010).
- Grandal-D'Anglade, A., Vidal, J. R. & Romaní A population study on the cave bear (*Ursus spelaeus* Ros.-Hein.) from Cova Eirós (Triacastela, Galicia, Spain). *Geobios* **30**, 723–731 (1997).
- Argant, A. Biochronologie et grands mammifères au Pléistocène moyen et supérieur en Europe occidentale: L'apport des canidés, des ursidés et des carnivores en général. *Quaternaire Revue de l'Association française pour l'étude du Quaternaire*. 467–480. <https://doi.org/10.4000/quaternaire.5334> (2009).
- Baryshnikov, G. F. & Puzachenko, A. Y. Craniometrical variability in the cave bears (Carnivora, Ursidae): Multivariate comparative analysis. *Quatern. Int.* **245**, 350–368 (2011).
- Santos, E., Gómez-Olivencia, A., Arlegi, M. & Arsuaga, J. L. Cranial morphological differences within *U. deningeri* – *U. spelaeus* lineage: A double traditional and geometric morphometrics approach. *Quatern. Int.* **433**, 347–362 (2017).
- Wagner, J. & Cermak, S. Revision of the early Middle Pleistocene bears (Ursidae, Mammalia) of Central Europe, with special respect to possible co-occurrence of spelaeoid and arctoid lineages. *Bull. Geosci.* **87**, 461–496 (2012).
- Barlow, A. et al. Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570 (2018).
- Knapp, M. From a molecule's perspective – contributions of ancient DNA research to understanding cave bear biology. *Hist. Biol.* **31**, 442–447 (2019).
- Valdiosera, C. et al. Typing single polymorphic nucleotides in mitochondrial DNA as a way to access Middle Pleistocene DNA. *Biol. Lett.* **2**, 601–603 (2006).
- Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.* **110**, 15758–15763 (2013).

22. Arsuaga, J. L. et al. How the Sima de los Huesos was won. *Anat. Rec.* **307**, 2225–2245 (2024).
23. García, N. & Arsuaga, J. L. Les carnivores (Mammalia) des sites du Pléistocène ancien et moyen d'Atapuerca (Espagne). *L'Anthropologie* **105**, 83–93 (2001).
24. Rabeder, G., Pacher, M. & Withalm, G. *Early Pleistocene Bear Remains from Deutsch-Altenburg (Lower Austria)* (ÖAW, Verlag der Österreichischen Akademie der Wissenschaften, 2010).
25. Wagner, J. Pliocene to early Middle Pleistocene ursine bears in Europe: A taxonomic overview. *J. Natl. Museum (Prague) Nat. History Ser.* **179**, 197–215 (2010).
26. Demarchi, B. et al. Protein sequences bound to mineral surfaces persist into deep time. *eLife* **5**, e17092 (2016).
27. Paterson, R. S. et al. Phylogenetically informative proteins from an Early Miocene rhinocerotid. *Nature* **643**, 719–724 (2025).
28. Taurozzi, A. J. et al. Deep-time phylogenetic inference by paleoproteomic analysis of dental enamel. *Nat. Protoc.* **19**, 2085–2116 (2024).
29. Cappellini, E. et al. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature* **574**, 103–107 (2019).
30. Welker, F. et al. The dental proteome of *Homo antecessor*. *Nature* **580**, 235–238 (2020).
31. Welker, F. et al. Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature* **576**, 262–265 (2019).
32. Madupe, P. P. et al. Enamel proteins reveal biological sex and genetic variability in southern African *Paranthropus*. *Science* **388**, 969–973 (2025).
33. Delsuc, F., Gasse, B. & Sire, J. Y. Evolutionary analysis of selective constraints identifies ameloblastin (AMBN) as a potential candidate for amelogenesis imperfecta. *BMC Evol. Biol.* **15**, 148 (2015).
34. Green, D. R. et al. Mapping the tooth enamel proteome and amelogenin phosphorylation onto mineralizing porcine tooth crowns. *Front Physiol* **10** (2019).
35. Gelse, K., Pöschl, E. & Aigner, T. Collagens—structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55**, 1531–1546 (2003).
36. García-Vázquez, A., Pinto-Llona, A. C., Maroto, J., Torres, T. & Grandal-D'Anglade, A. Characterising the cave bear *Ursus spelaeus* Rosenmüller by ZooMS: a review of peptide mass fingerprinting markers. *Earth Environ. Sci. Trans. Royal Soc. Edinb.* **114**, 83–93 (2023).
37. Asaka, T. et al. Type XVII collagen is a key player in tooth enamel formation. *Am. J. Pathol.* **174**, 91–100 (2009).
38. Goettig, P., Magdolen, V. & Brandstetter, H. Natural and synthetic inhibitors of kallikrein-related peptidases (KLKs). *Biochimie* **92**, 1546–1567 (2010).
39. Sawafuji, R. et al. Proteomic profiling of archaeological human bone. *Royal Soc. Open. Sci.* **4**, 161004 (2017).
40. Friedrich, A. K. U. et al. Comparative coagulation studies in hibernating and summer-active black bears (*Ursus americanus*). *Thromb. Res.* **158**, 16–18 (2017).
41. Chow, B. A., Donahue, S. W., Vaughan, M. R., McConkey, B. & Vijayan, M. M. Serum immune-related proteins are differentially expressed during hibernation in the American black bear. *PLOS ONE* **8**, e66119 (2013).
42. Welinder, K. G. et al. Biochemical foundations of health and energy conservation in hibernating free-ranging subadult brown bear *Ursus arctos*. *J. Biol. Chem.* **291**, 22509–22523 (2016).
43. Mugahid, D. A. et al. Proteomic and transcriptomic changes in hibernating grizzly bears reveal metabolic and signaling pathways that protect against muscle atrophy. *Sci. Rep.* **9**, 19976 (2019).
44. Takamatsu, N. et al. Expression of multiple α 1-antitrypsin-like genes in hibernating species of the squirrel family. *Gene* **204**, 127–132 (1997).
45. Kondo, N. & Kondo, J. Identification of novel blood proteins specific for mammalian hibernation. *J. Biol. Chem.* **267**, 473–478 (1992).
46. Cooper, S. et al. Platelet proteome dynamics in hibernating 13-lined ground squirrels. *Physiol. Genom.* **53**, 473–485 (2021).
47. Gil-Bona, A. & Bidlack, F. B. Tooth enamel and its dynamic protein matrix. *Int. J. Mol. Sci.* **21**, 4458 (2020).
48. Froment, C. et al. Protein sequence comparison of human and non-human primate tooth proteomes. *J. Proteom.* **231**, 104045 (2021).
49. Noonan, J. P. et al. Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597–599 (2005).
50. Fong-Zazueta, R. et al. Phylogenetic signal in primate tooth enamel proteins and its relevance for paleoproteomics. *Genome Biol. Evol.* **17**, evaf007 (2025).
51. van Heteren, A. H., Arlegi, M., Santos, E., Arsuaga, J. L. & Gómez-Olivencia, A. Cranial and mandibular morphology of Middle Pleistocene cave bears (*Ursus deningeri*): implications for diet and evolution. *Hist. Biol.* **31**, 485–499 (2019).
52. Parés, J. M. & Pérez-González, A. Paleomagnetic Age for Hominid Fossils at Atapuerca Archaeological Site, Spain. *Science* **269**, 830–832 (1995).
53. Rodríguez, J. et al. One million years of cultural evolution in a stable environment at Atapuerca (Burgos, Spain). *Q. Sci. Rev.* **30**, 1396–1412 (2011).
54. Cuenca-Bescós, G. et al. Comparing two different Early Pleistocene microfaunal sequences from the caves of Atapuerca, Sima del Elefante and Gran Dolina (Spain): Biochronological implications and significance of the Jaramillo subchron. *Quatern. Int.* **389**, 148–158 (2015).
55. van der Made, J., Rosell, J. & Blasco, R. Faunas from Atapuerca at the Early–Middle Pleistocene limit: The ungulates from level TD8 in the context of climatic change. *Quatern. Int.* **433**, 296–346 (2017).
56. Stefen, C. Enamel structure of arctoid carnivora: Amphicyonidae, Ursidae, Procyonidae, and Mustelidae. *J. Mammal.* **82**, 450–462 (2001).
57. Wiszniowska, T. et al. Dental enamel structure in fossil bears *Ursus spelaeus* and *U. wenzensis* (=minimus) in comparison to selected representatives of other Carnivora. Morphology and systematics of fossil vertebrates. DN Publisher, Wrocław 125 (2010).
58. Mackiewicz, P. et al. Analysis of dental enamel thickness in bears with special attention to *Ursus spelaeus* and *U. wenzensis* (=minimus) in comparison to selected representatives of mammals in (2010). (DN Publisher).
59. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
60. Chiva, C. et al. QCloud: A cloud-based quality control system for mass spectrometry-based proteomics laboratories. *PloS one*. **13**, e0189209 (2018).
61. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
62. Gutenbrunner, P., Kyriakidou, P., Welker, F. & Cox, J. Spectrum graph-based de-novo sequencing algorithm MaxNovo achieves high peptide identification rates in collisional dissociation MS/MS spectra. (2021). bioRxiv 2021–09.
63. Han, Y., Ma, B. & Zhang, K. SPIDER: software for protein identification from sequence tags with de novo sequencing error in Proceedings. IEEE Computational Systems Bioinformatics Conference, 2004. CSB., (2004), pp. 206–215., (2004), pp. 206–215.
64. Zhang, J. et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cell. proteomics* **11** (2012).
65. Madeira, F. et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
66. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

67. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
68. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
69. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*. **14**, 587–589 (2017).
70. Ronquist, F. et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
71. Perez-Riverol, Y. et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553 (2025).

Acknowledgements

We thank Johanna Krueger, Joanna L. Kelley, Guillermo Carrillo Martín and Luis Ferrández Peral for their advice during planning and data analysis.

Author contributions

Conceptualization: M.M.-T. and E.S.; Data curation: A.G.-C.; Formal analysis: A.G.-C., Methodology and analysis: A.G.-C., with assistance and guidance from M.M.-T., E.S., T.M.-B., and E.L., and with specific contributions from R.F.-Z.; Paleontological material contribution: E.S., D.M.A., A.G.-O., A.G.-A., L.R., L.P., M.S., A.G.-V., M.VdA. and J.L.A.; Writing original draft: A.G.-C.; Writing-review and editing: A.G.-C. with input from all coauthors; Supervision: M.M.-T., E.S., and T.M.-B.; Project administration and funding acquisition: J.L.A., J.M.BdC and M.M.-T.

Funding

A.G.-C. was supported by the Horizon 2020 research and innovation program of the European Union under the Marie Skłodowska-Curie European Training Network (ETN) “PUSHH” (grant agreement No. 861389), and Atapuerca Foundation predoctoral grant, supported by Reale Foundation. E.S. has an Atapuerca Foundation postdoctoral grant. M.M.-T. receives funding from The Leakey Foundation through the personal support of W.D. Crook. E.L., T.M.-B., and D.M.A. acknowledge the support of the CERCA Programme/Generalitat de Catalunya. Fieldwork at the Atapuerca sites is funded by the Junta de Castilla y León and the Fundación Atapuerca. The Atapuerca research project is financed by the Ministerio de Ciencia, Innovación y Universidades Grant PID2021-122355NB-C31 and C33, financed by MCIN/AEI/10.13039/501100011033/FEDER, EU, and , PID2024-156477NB-C32 financed by MICIU/AEI/10.13039/501100011033/FEDER, UE. The Baio project is funded by Diputación Foral de Gipuzkoa, Research Group IT1485-22 from the Eusko Jaurlaritza-Gobierno Vasco, the grant CNS2023-143739 funded by MICIU/AEI/10.13039/501100011033/FEDER, EU, and by European Union NextGenerationEU/PRTR. Fieldwork at Aroeira site is funded by PALAEO.WEST.IBERIA, funded through the Fundação para a Ciência e a Tecnologia (FCT), project number 2023.16301.ICDT / LIS-BOA20230-FEDER-00865300, Câmara Municipal de Torres Novas, RENOVÁ-Fábrica de Papel do Almonda, Sociedade Torrejana de Espeleologia e Arqueologia and Paleoalmonda. M.S. funded by Ramon y Cajal postdoctoral grant (RYC2021-032999-I) with financial sponsorship of Ministerio de Ciencia e Innovación and the European Union NextGenerationEU.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-58093-x>.

Correspondence and requests for materials should be addressed to A.G.-C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026