



OPEN

Comparing extraction method efficiency for high-throughput palaeoproteomic bone species identification

Dorothea Mylopotamitaki^{1,2}✉, Florian S. Harking³, Alberto J. Taurozzi⁴, Zandra Fagnäs⁴, Ricardo M. Godinho⁵, Geoff M. Smith^{2,6}, Marcel Weiss^{2,7}, Tim Schüler⁸, Shannon P. McPherron⁹, Harald Meller¹⁰, João Cascalheira⁵, Nuno Bicho⁵, Jesper V. Olsen³, Jean-Jacques Hublin^{1,2} & Frido Welker⁴✉

High-throughput proteomic analysis of archaeological skeletal remains provides information about past fauna community compositions and species dispersals in time and space. Archaeological skeletal remains are a finite resource, however, and therefore it becomes relevant to optimize methods of skeletal proteome extraction. Ancient proteins in bone specimens can be highly degraded and consequently, extraction methods for well-preserved or modern bone might be unsuitable for the processing of highly degraded skeletal proteomes. In this study, we compared six proteomic extraction methods on Late Pleistocene remains with variable levels of proteome preservation. We tested the accuracy of species identification, protein sequence coverage, deamidation, and the number of post-translational modifications per method. We find striking differences in obtained proteome complexity and sequence coverage, highlighting that simple acid-insoluble proteome extraction methods perform better in highly degraded contexts. For well-preserved specimens, the approach using EDTA demineralization and protease-mix proteolysis yielded a higher number of identified peptides. The protocols presented here allowed protein extraction from ancient bone with a minimum number of working steps and equipment and yielded protein extracts within three working days. We expect further development along this route to benefit large-scale screening applications of relevance to archaeological and human evolution research.

The majority of the archaeological skeletal record is dominated by bone specimens that cannot be assigned a species identification based solely on morphological characteristics. As a result, the increasing application of bio-molecular methods over the past two decades in archaeology, palaeoanthropology, and paleontology has seen the development of genetic^{1,2} and proteomic^{3–5} methods for high-throughput taxonomic identification approaches of such bone assemblages. Proteomically, these include the development of MALDI-TOF MS-based peptide mass fingerprinting (PMF) of collagen type I⁴, data-independent acquisition (DIA)-based species identification based on limited bone proteome sequence databases³, or the in-depth characterization of entire skeletal proteomes^{5–7}.

In particular, the application of collagen PMF (also known as Zooarchaeology by Mass Spectrometry, or ZooMS) to skeletal remains has found widespread adoption in the proteomic screening of skeletal assemblages in archaeological contexts. Studies have demonstrated that large-scale screening of thousands of bone fragments is feasible^{8,9} in different geographic regions across the globe. Such studies have provided information on

¹Chaire de Paléoanthropologie, CIRB (UMR 7241–U1050), Collège de France, Paris, France. ²Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ³Center for Protein Research, University of Copenhagen, Copenhagen, Denmark. ⁴Globe Institute, University of Copenhagen, Copenhagen, Denmark. ⁵Interdisciplinary Center for Archaeology and Evolution of Human Behaviour, University of Algarve, Faro, Portugal. ⁶School of Anthropology and Conservation, University of Kent, Kent, UK. ⁷Institut für Ur- und Frühgeschichte, Friedrich-Alexander-Universität, Erlangen, Germany. ⁸Thuringian State Office for the Preservation of Historical Monuments and Archaeology, Weimar, Germany. ⁹Department of Human Origins, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ¹⁰State Office for Heritage Management and Archaeology, Saxony-Anhalt—State Museum of Prehistory, Halle (Saale), Germany. ✉email: dorothea.mylopotamitaki@palaeoproteomics.org; frido.welker@sund.ku.dk

the ecological context in which hominins, including humans, operated^{10–12}, provided new insights into hunting strategies, animal resource processing, or herd management strategies^{13–16}, and the selection of particular animal taxa for the production of bone tools^{17–21}. Despite its widespread adoption in the archaeological community, collagen-based PMF suffers from comparatively low taxonomic resolution and an absence of a widely adopted, computational approach to spectral identification, preventing further adoption of this approach in the wider research community.

Liquid-chromatography tandem mass spectrometry (LC–MS/MS) overcomes or minimizes the challenges associated with MALDI-TOF MS. Some of the main challenges in MALDI-TOF spectra are the manual assignment of peptide peaks for species identification that is time-consuming, the low mass resolution, and the absence of directly obtained amino acid sequence information. Together this results in taxonomic identifications that are less precise as what would, at least theoretically, be possible using LC–MS/MS analysis of the same proteomes³. LC–MS/MS allows the generation of MS2 spectra for which a peptide sequence is determined using a computational search program, e.g. MaxQuant²². From this information, the presence of specific proteins, or protein groups, are inferred. In addition, the peptide sequences can form the basis of taxonomic or phylogenetic analysis. The improvement in accuracy, speed, and sensitivity of the last generation of MS instrumentation²³, together with the continuous growth of protein databases²⁴, allows the identification of thousands of unique proteins for each MS run from low input amounts²⁵. LC–MS/MS-based approaches are therefore suited to study the highly-degraded, low-quantity proteomes preserved in archaeological remains. LC–MS/MS-based studies of archaeological and anthropological materials initially focused on collagen type I, the dominant bone protein, but increasingly include entire ancient proteomes^{26–30}. Despite the relatively slow rate of protein single amino acid polymorphism (SAP) accumulation, compared to nucleotide variation accumulated at the genetic level, LC–MS/MS analysis has allowed the phylogenetic analysis of ancient protein datasets for a range of animal taxa^{26,29,31–35}.

Due to the experimental simplicity of mass spectrometry-based proteomics, and the capacity to process large cohorts of samples, label-free quantification approaches are most frequently used. Current MS/MS approaches largely rely on data-dependent acquisition (DDA) for precursor ion selection, according to their abundances. This maximizes the success of peptide sequence determination but limits reproducibility and quantitative potential. To resolve these issues, in recent years, several data-independent acquisition (DIA) mass spectrometric strategies, including SWATH-MS³⁶ (sequential window acquisition of all theoretical fragment ion spectra), HDMSE³⁷ (high definition MSE), and AIF³⁸ (all-ion fragmentation), were established. DIA implements a parallel fragmentation of all precursor ions, regardless of their intensity or other characteristics, thereby enabling the establishment of a complete record of a sample³⁶. DIA approaches are now in development for ancient protein analysis and offers the potential to extend the dynamic range of MS/MS data acquisition by generating data from more peptides, especially lower abundance peptides, while also improving reproducibility and quantification.

Resulting from these developments, the “Species by Proteome INvestigation” (SPIN) is a recently proposed proteomics workflow leveraging automatic approaches to LC–MS/MS data analysis in association with shorter liquid chromatography separation and DIA or DDA spectral acquisition. SPIN was proposed with a single-step protein extraction method from mineralized tissues followed by digestion using protein aggregation capture (PAC)^{39,40}. Shortening the LC–MS/MS analysis to less than 10 min became possible with new LC technology⁴¹ and fast-scanning data-dependent or multiplexed data-independent tandem MS acquisition methods²³. SPIN employs an automated approach to achieve a taxonomic assignment, based on protein sequence databases with gene-wise alignments. Although demonstrated to be successful in Late Pleistocene (LP) and Holocene archaeological settings, initial results show that SPIN has a comparatively low success rate for some archaeological sites³.

To further explore the high-throughput capacities of SPIN proteomic analyses, we designed a comparison of six protein extraction approaches (Fig. 1). The selected extraction methods are commonly used for proteomic extractions from archaeological bone specimens^{4,5,42–46} and allow an easy scale-up for the processing of hundreds of specimens simultaneously. We apply these extraction approaches to 12 bone specimens from two Late Pleistocene (LP) cave sites with different preservation: the site Ilsenhöhle Ranis (50°39.7563'N, 11°33.9139'E, Germany, hereafter: Ranis), and Gruta da Companheira (N 37°09.19'N, 8°31.47'W, Portugal, hereafter: GdC or Companheira). Skeletal remains from LP sites are usually highly degraded and fragmented which prevents morphological species identification. Specimens from Companheira were selected as previous SPIN research at GdC indicated variable and challenging proteome preservation at the site³. In contrast, ongoing research at Ranis indicates excellent molecular preservation (unpublished data). This allowed us to assess the performance of the different proteome extraction methods in terms of proteome complexity, protein sequence coverage, and accuracy of species identification for two LP sites with different extents of proteomic preservation.

Results

We analyzed 12 different bone specimens using six extraction methods in total (Fig. 1) and injected 10% of the resulting peptide digestion for LC–MS/MS analysis using an EvoSep One instrument coupled to an Exploris 480 orbitrap mass spectrometer. Six of the bone specimens are well-preserved bone specimens deriving from Ranis, Germany, and the remaining six specimens are degraded bone specimens deriving from GdC, Portugal (Table 1 and Supplementary Fig S1, SI). For extraction methods 3a and 5a, we were only able to analyze 4 specimens, including 2 blanks, due to column clogging either due to EDTA precipitation in acidic conditions or overloaded Evotips. Additionally, method 3b generated no reliable results for Ranis specimens. Consequently, methods 3a, 5a and 3b-Ranis were excluded from this study.

All generated data were analysed in MaxQuant⁴⁷ (v. 2.1.1.0) in both “specific” mode, where trypsin-specific cleavage is required at both peptide termini, and semi-specific” mode, where non-specific cleavage is allowed at one peptide terminus (see ‘Methods- MaxQuant search’). As no MS spectra were generated for both extraction methods 3a and 5a, they were excluded from the study. In addition, for the Ranis specimens, we did not

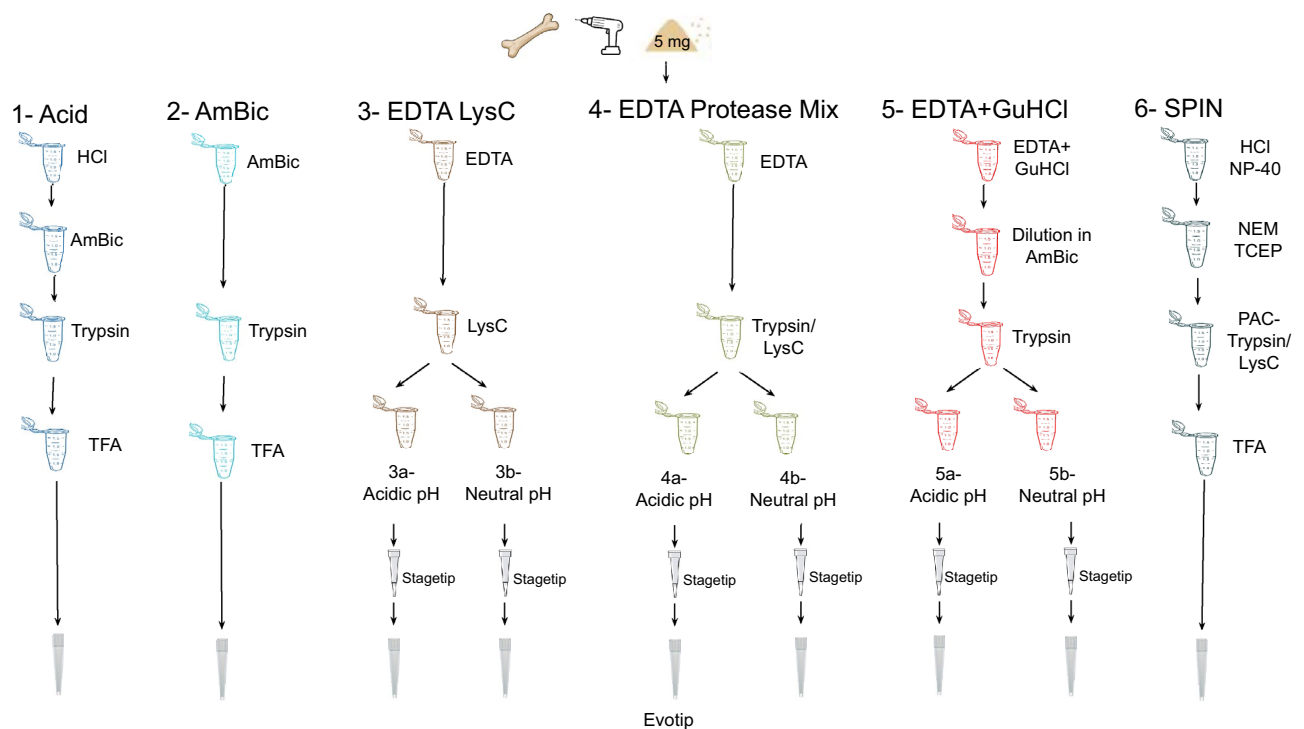


Figure 1. Schematic overview of extraction conditions of the six proteomic extraction methods compared in this study. The extraction approaches 3a and 5a generated no MS spectra and were excluded from further analysis.

| Specimen number | Site | Layer | Chronological age | Taxonomic identity |
|-----------------|--------------------------------|--------------------|-------------------|-----------------------------------|
| GdC-1 | Gruta da Companheira, Portugal | Galeria 1 | > 50 kya | Ursidae |
| GdC-6 | Gruta da Companheira, Portugal | Galeris 2- Level 2 | > 50 kya | Ursidae/ Capra/ Bovidae/ Reindeer |
| GdC-7 | Gruta da Companheira, Portugal | Galeria 2 | > 50 kya | Not previously identified |
| GdC-9 | Gruta da Companheira, Portugal | Galeria 2 | > 50 kya | Reindeer |
| GdC-26 | Gruta da Companheira, Portugal | Galeria 2 | > 50 kya | Bovidae |
| GdC-193 | Gruta da Companheira, Portugal | Galeria 2 | > 50 kya | Not previously identified |
| R10300 | Ilsehöhle Ranis, Germany | X (Graue Schicht) | 40–50 kya | Rhinocerotidae |
| R10329 | Ilsehöhle Ranis, Germany | X (Graue Schicht) | 40–50 kya | Reindeer |
| R10340 | Ilsehöhle Ranis, Germany | X (Graue Schicht) | 40–50 kya | Ursidae |
| R10357 | Ilsehöhle Ranis, Germany | X (Graue Schicht) | 40–50 kya | Equidae |
| R10359 | Ilsehöhle Ranis, Germany | X (Graue Schicht) | 40–50 kya | Hyaenidae/Felinae |
| R10363 | Ilsehöhle Ranis, Germany | X (Graue Schicht) | 40–50 kya | Reindeer |

Table 1. Bone specimens used in this study. The chronological age indicated is approximate and based on estimates in the archaeological literature. When available, taxonomic identity is based on prior analysis using ZooMS (for Ranis) or SPIN (for GdC).

generate reliable data with extraction approach 3b. We find that the “specific” search allowed the identification of approximately 250 additional MS2 spectra for Ranis compared to a “semi-specific” search, but not for GdC, while all other analyses provided identical insights (Supplementary Fig. S2, SI). Therefore, we present data on the “specific” searches below, with the comparative results of the “semi-specific” search provided in the Supplementary Information.

Spectral acquisition

Optimized LC–MS/MS analysis aims to strike a balance between MS1 cycle times and MS2 spectral acquisition rates. We observed that, regardless of extraction method, the GdC samples generated more acquired MS1 scans than Ranis ($F = 153.68$, $p < 2.2e - 16$; Fig. 2a). For Ranis, extraction approach 6 generated approximately 2500–4500 MS1 scans, while all other extraction methods obtained approximately 2500 MS1 scans each. For GdC, extraction approach 4a generated the highest number of MS1 scans (> 6000) while the remaining extraction

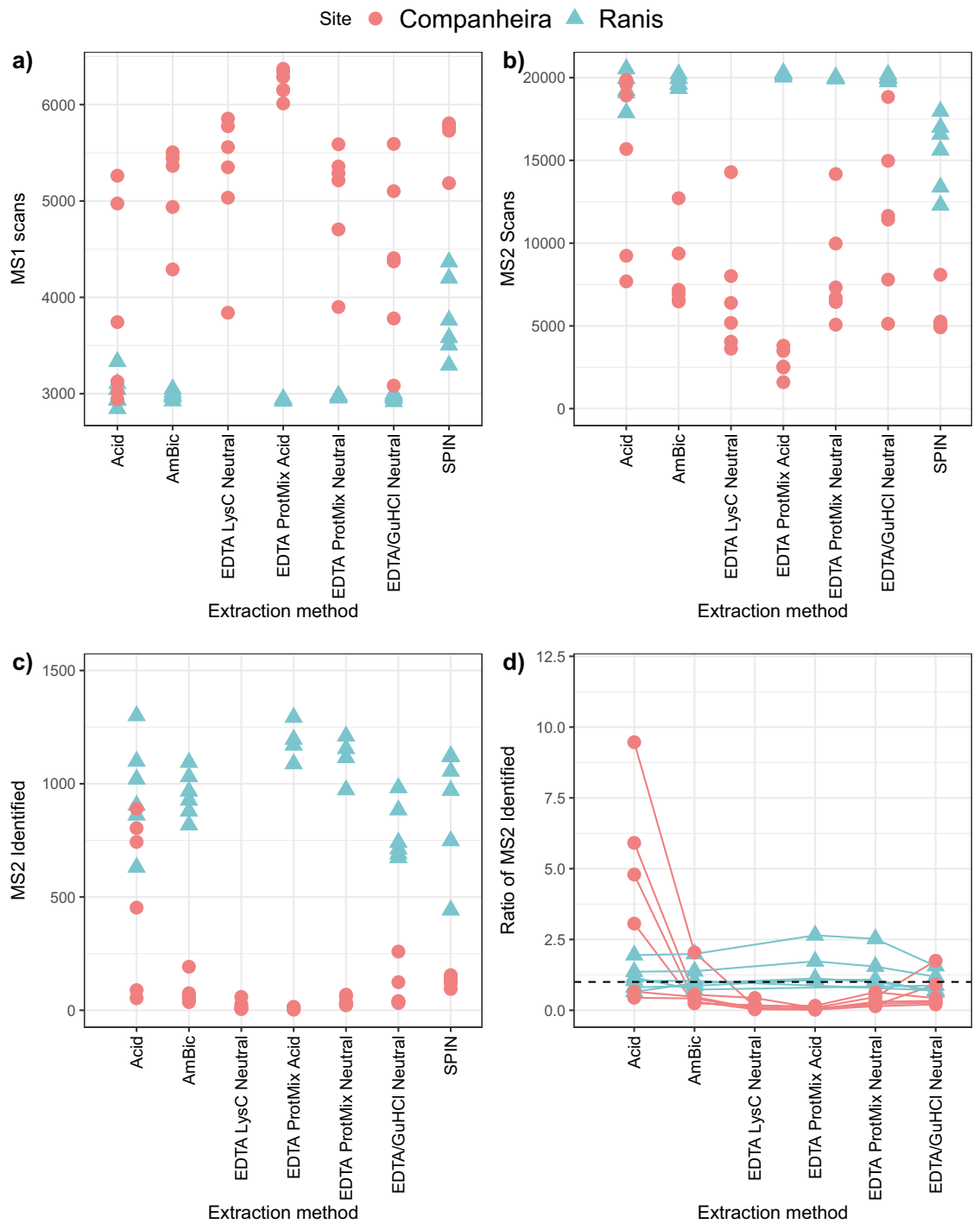


Figure 2. Summary information on MS data acquisition in the “specific” MaxQuant search mode, by extraction method; 1-Acid, 2-AmBic, 3b-EDTA LysC Neutral, 4a-EDTA Protease Mix Acidic, 4b-EDTA Protease Mix Neutral, 5b-EDTA + GuHCl Neutral, 6-SPIN. (a) The number of MS spectra recorded in each raw file per specimen, (b) the number of MS/MS spectra recorded in the raw files acquired per specimen, (c) the total number of identified tandem MS spectra, and (d) the ratio of identified tandem MS spectra in comparison to SPIN for each specimen. In (d), the dashed line represents extraction method 6 (value = 1). Extraction methods 3a-EDTA LysC Acidic and 5a-EDTA + GuHCl Acidic were excluded from the study as no reliable MS spectra were generated.

methods acquired approximately 2500–6000 MS1 scans, with generally a high variability between specimens ($F = 6.95$, $p < 9.452e - 06$). These observations are evidence of the generally lower proteome preservation in the GdC specimens.

In contrast to MS1, a reverse pattern was obtained for MS2 spectral acquisition (Fig. 2b). Here, the acquired MS2 scan numbers were higher for the Ranis specimens than GdC specimens ($F = 161.19$, $p < 2e - 16$). For Ranis specimens, all extraction methods except method 6 resulted in the acquisition of approximately 20,000 MS2 scans. Method 6 resulted in 12,000–17,000 MS2 spectra. For GdC specimens, MS2 acquisition varied. The highest number of MS2 scans (approximately 5500–20,000) was obtained with extraction approach 1, while method 5b generated approximately 5000–18,000 MS2 scans ($F = 6.60$, $p < 1.7e - 05$).

We found clear differences in MS2 spectral identification between Ranis and GdC bone extracts for all extraction methods ($F = 307.73$, $p < 2e - 16$) (Fig. 2c). These observations also extend to the “semi-specific” search results (Supplementary Fig. S3, SI). There was an overlap in identified MS2 counts between extraction methods. We found that extraction approach 1 generated the highest number of identified MS2 scans for Ranis bone specimens (approximately 500–1500). Identified MS2 scans with extraction method 4, both in 4a and 4b, showed a higher consistency among Ranis specimens (> 900 MS2 identified scans). For GdC specimens, extraction protocol 1 was the only approach resulting in over 500 identified MS2 scans (for four out of six specimens from GdC). All the remaining proteomic extraction methods recovered less than 300 identified MS2 scans for GdC specimens. We also observed significant differences for the average ion intensity among all the extraction methods for both archaeological sites (Supplementary Fig. S4, SI) in the “specific” ($F = 596.09$, $p < 2.2e - 16$) and in the “semi-specific” search ($F = 454.58$, $p < 2.2e - 16$).

Finally, we compared the number of identified MS2 scans obtained with extraction method 6 to the other extraction methods ($F = 3.08$, $p = 0.010$; Fig. 2d). For Ranis specimens, we observed that extraction method 4, in both 4a and 4b conditions, generated more identified MS2 scans for several specimens compared to method 6. The remaining extraction approaches showed no significant differences in the identified MS2 scans compared to method 6. For GdC specimens, extraction protocol 1 obtained a higher ratio of identified MS2 spectra compared to method 6, while all other approaches resulted in a lower MS2 identification ratio.

Proteome composition and species identification

The SPIN identification approach³ focuses on 20 protein-coding genes and uses gene-wise sequence coverage estimation for species identification. To further investigate the efficiency of extraction methods to recover peptides belonging to non-collagenous proteins (NCPs), we compared the protein sequence coverage (site counts)³ for each specimen. Most of the covered sequences were concentrated in the two collagen type I chains for all the specimens. The number of identified amino acids decreased in all GdC specimens and the majority of the methods generated no amino acid sequence coverage of non-collagenous proteins for GdC specimens (Fig. 3 and Supplementary Table S1, SI). Extraction approach 1 generated a high number of amino acid site counts for most of the GdC specimens, while extraction approaches 4b and 6 had the highest site counts for Ranis specimens. Non-matching species-specific site counts (amino acid sequences for different taxa or closely related species) were below 500 apart from one specimen with approach 3b (not shown). In agreement with the site counts, extraction methods 4b and 6 obtained the highest number of gene counts, NCP gene counts, unique peptides, and NCP site count for Ranis specimens, while extraction approach 1 showed the best results for the GdC specimens.

We also compared the species identification provided for the same specimen per extraction method (Fig. 4 and Supplementary Table S1, SI). Ranis specimens generated compatible species identification for all of the extraction methods apart from extraction approach 3b, which gave no data for all Ranis specimens. In general, most of GdC specimens were not identified; only two specimens (Gdc_1 and Gdc_9) were correctly assigned to the same species as previously identified, while Gdc_6 was falsely assigned with extraction approach 1.

Peptide characteristics

Over time, ancient proteomes are modified in various ways due to diagenesis, such as protein deamidation and peptide bond cleavage⁴⁸. Therefore, we explored differences in the estimated extent of deamidation and semi-specific cleavage rates in our extracts to quantify the effect different extraction protocols may have on these variables. As all extraction methods were performed on the same homogenized bone powders, there should be no significant differences in the obtained values between the extraction methods. Our analysis of deamidation rates was complicated due to the low-quality spectra generated for GdC specimens ($F = 2.8206$, $p = 0.01746$). For Ranis specimens, we observed generally consistent rates of glutamine and asparagine deamidation across the neutral extraction methods and slightly higher deamidation values for acidic demineralization methods (Fig. 5 and Supplementary Fig. S5, SI, excluding method 6).

Mis-cleaved peptides for each extraction method per site were also estimated to validate our results for both searches (specific and semi-specific). The specific search was set up to 2 mis-cleaved peptides while the semi-specific search was up to 4 mis-cleavages for both archaeological sites by MaxQuant default parameters (Supplementary Figs. S7 and S8, SI). Extraction approach 3a is excluded from this comparison because LysC only cleaves at lysine (K) sites of the protein sequence. In the specific search, methods 5b and 1 had the highest rate of mis-cleavages for Ranis and GdC specimens accordingly, while method 6 had the lowest for both sample sets apart from extraction method 4a. In the semi-specific search, methods 5b and 6 had the highest rate of mis-cleavages for Ranis and GdC specimens accordingly.

The distribution of peptides according to their hydrophobicity (as expressed by the grand average of hydropathy GRAVY) scale using the method of Kyte and Doolittle⁴⁹ was evaluated for all proteomic extraction methods (Supplementary Fig. S9, SI). The average GRAVY scores for all extraction methods are below 0, indicating a higher number of hydrophilic peptides. The hydrophobicity scores for Ranis specimens were not significantly

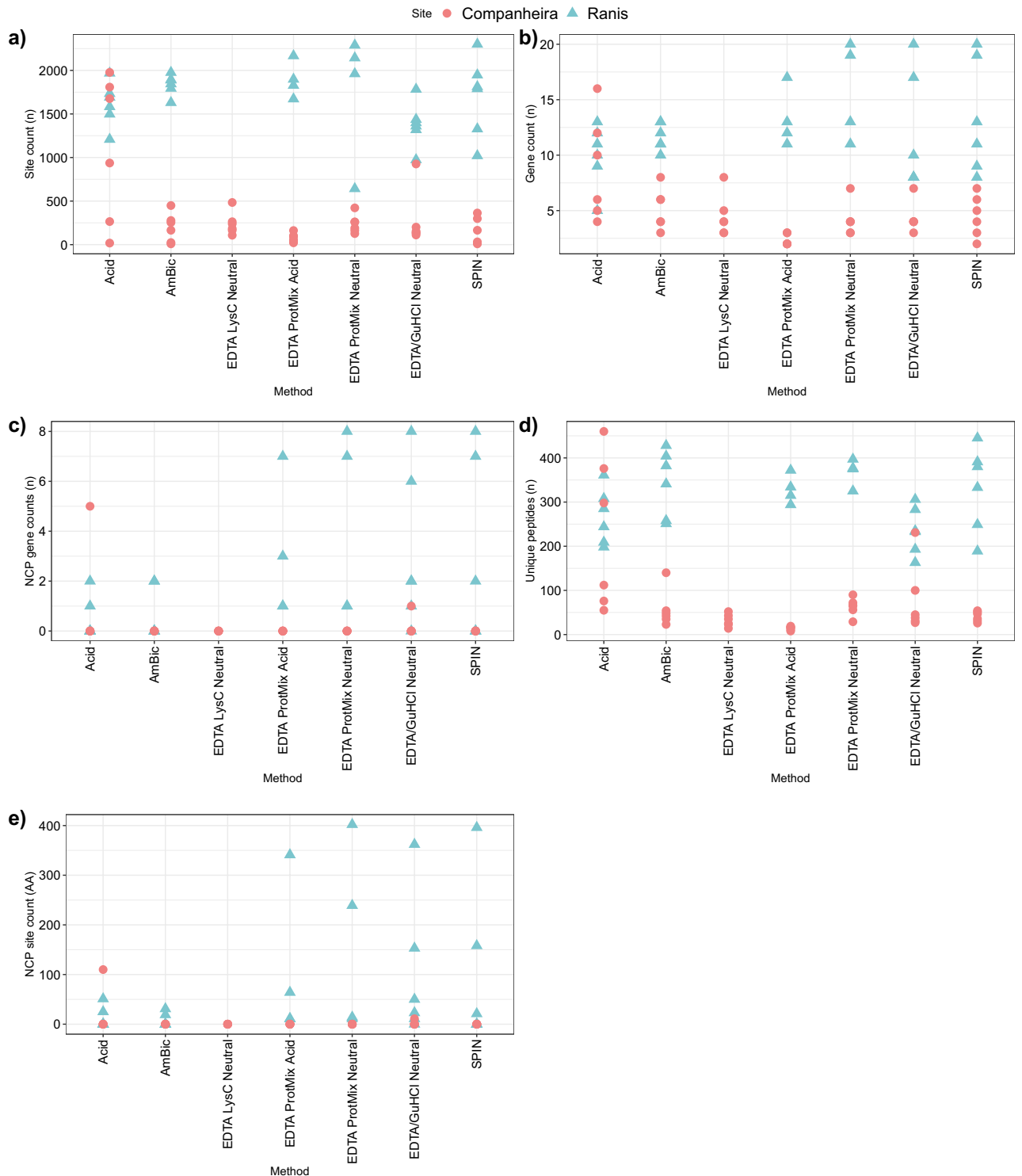


Figure 3. Proteome composition and peptide recovery for all the specimens by extraction method in the “specific” MaxQuant search; 1-Acid, 2-AmBic, 3b-EDTA LysC Neutral, 4a-EDTA Protease Mix Acidic, 4b-EDTA Protease Mix Neutral, 5b-EDTA + GuHCl Neutral, 6-SPIN. **(a)** Number of amino acid positions identified for species identification. **(b)** Number of identified protein genes for each specimen. **(c)** Number of non-collagenous proteins identified for each specimen. **(d)** Number of unique peptide sequences recovered for each specimen. **(e)** Number of amino acids for non-collagenous proteins recovered for each specimen. Extraction methods 3a-EDTA LysC Acidic and 5a-EDTA + GuHCl Acidic were excluded from the study as no reliable MS spectra were generated.

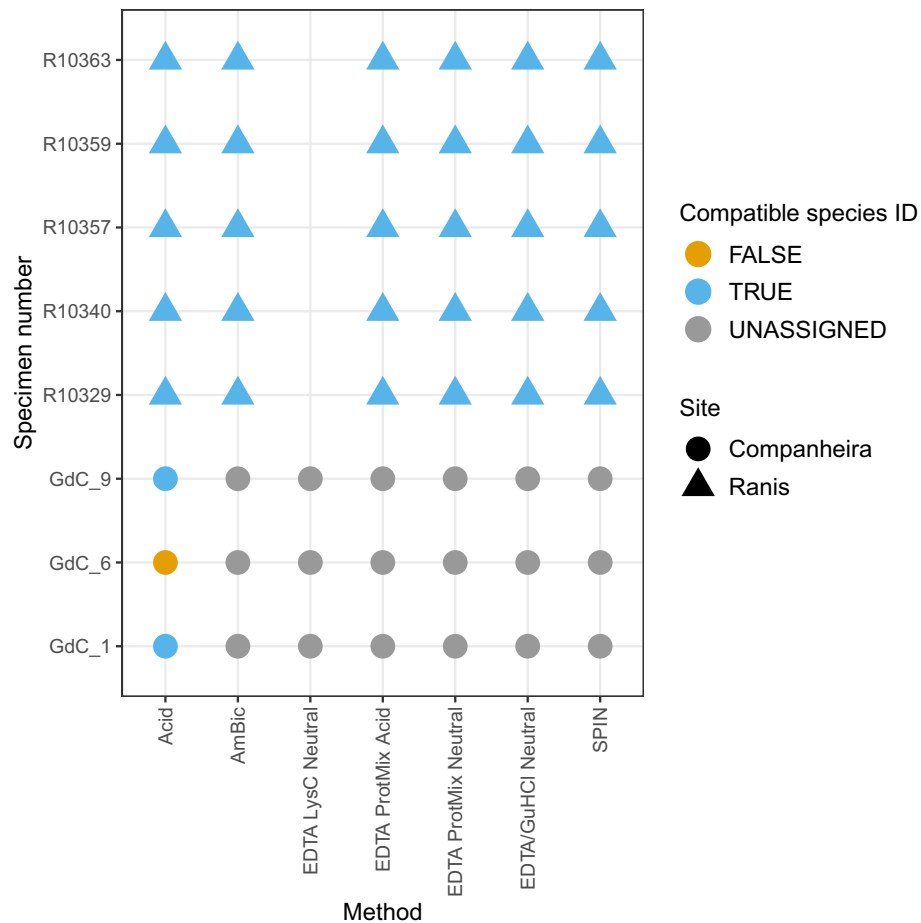


Figure 4. Compatibility of species identification in the “specific” MaxQuant search per specimen for each protein extraction approach; 1-Acid, 2-AmBic, 3b-EDTA LysC Neutral, 4a-EDTA Protease Mix Acidic, 4b-EDTA Protease Mix Neutral, 5b-EDTA + GuHCl Neutral, 6-SPIN. “Unassigned” specimens did not generate reliable results for species identification compared to their previous taxonomic identities. Of the 12 analyzed specimens in this study, specimens “GdC_7”, “GdC_26” and “GdC_193” were excluded from this table because there was no precise species identification assigned previously. “Ra_10300” was also excluded from the table because no member of the Rhinocerotidae family is included in the Rütther et al. 2022³ reference protein sequence database. Extraction methods 3a-EDTA LysC Acidic and 5a-EDTA + GuHCl Acidic were excluded from the study as no reliable MS spectra were generated.

different for all extraction methods, while GdC specimens showed different values of hydrophobicity depending on the extraction method ($F = 28.16$, $p = 1.247e - 06$). The extraction methods 1, 2, and 3b seemed to cluster together and 4a, 4b, 5b, and 6 were grouped together in a higher hydrophathy score ($F = 6.31$, $p = 2.421e - 05$).

Discussion and conclusion

Archaeological and palaeontological skeletal remains are a limited resource but contain an extensive amount of information about the past. The development of high-throughput shotgun proteomics (HTS) approaches allows us to identify skeletal remains of otherwise scarce animal species, such as hominins. Hence, optimizing methods of skeletal proteome extraction from small sample sizes with poor preservation becomes relevant. We, therefore, compared six proteomic extraction approaches on Late Pleistocene (LP) remains with variable proteomic preservation.

Different extraction approaches yielded the highest number of identified peptides based on the preservation condition of the specimens. Even though we were able to retrieve more species-specific amino acid sequences with the acidic extraction approach, we acquired valid species identifications for only two specimens from GdC. For Ranis specimens, extraction method 4a generated the highest number of identified peptides, which is in accordance with previous studies⁵⁰. In contrast, extraction method 1 outperformed the other methods for GdC samples. Even though an acidic extraction environment might negatively affect peptide fragmentation and MS2 intensity, we recovered a higher number of unique identified peptides and valid species identifications in acidic environments. This might be explained based on acidic demineralization solutions resulting in proteins being more effectively released from the bone skeletal matrix in association with the removal of protease inhibitors, in contrast with neutral solutions⁴⁶.

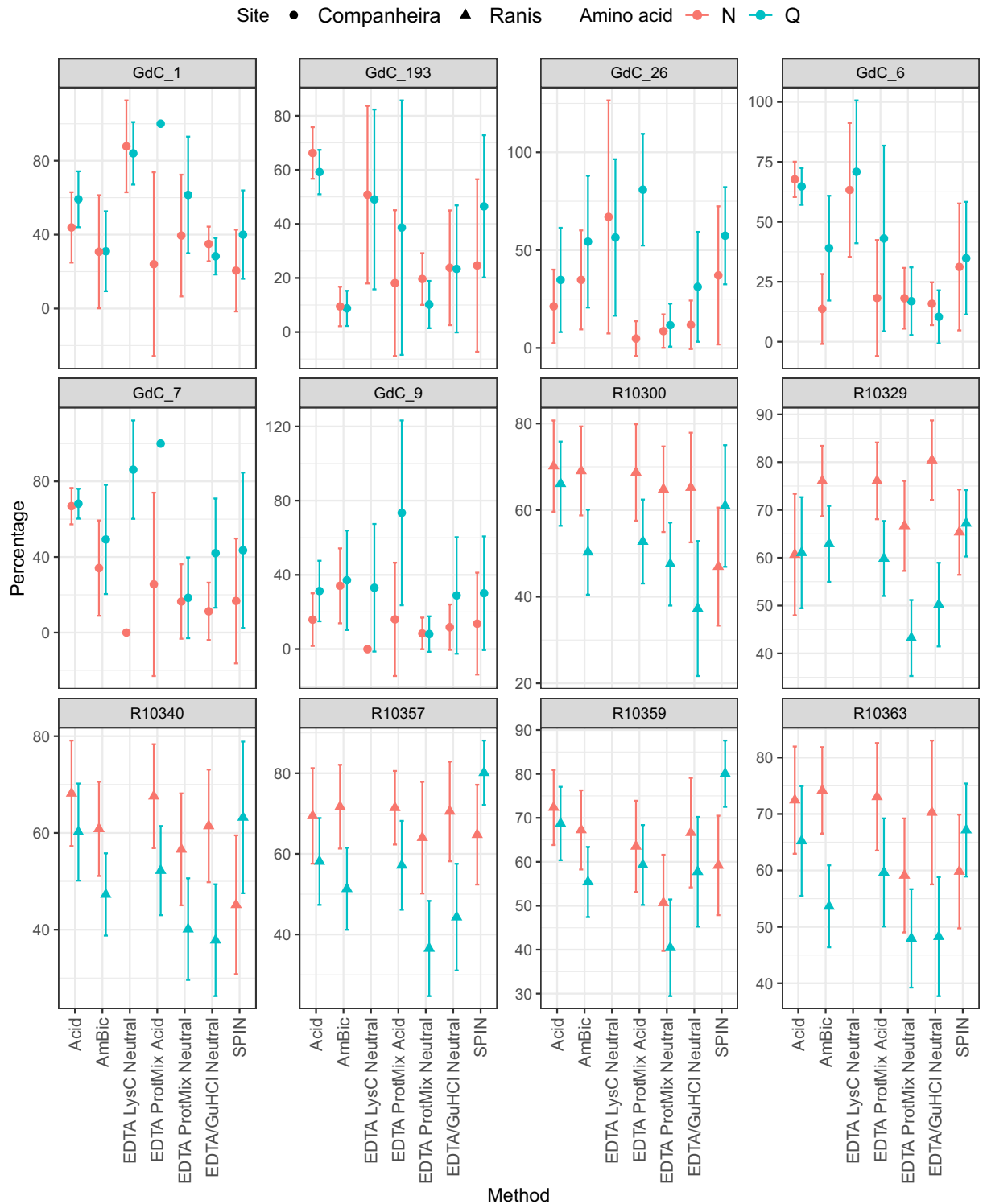


Figure 5. Amino acid deamidation in the “specific” MaxQuant search for each sample by extraction method; 1-Acid extraction, 2-AmBic, 3b-EDTA LysC Neutral, 4a-EDTA Protease Mix Acidic, 4b-EDTA Protease Mix Neutral, 5b-EDTA + GuHCl Neutral, 6-SPIN. Glutamine (Q) and asparagine (N) deamidation for each sample by extraction method. The y-axis represents the percentage of deamidation rate for Q and N, where 100% indicates complete deamidation and 0% indicates no deamidation. Error bars represent 2SD. EDTA LysC Neutral (3b) extraction method generated no results for Ranis specimens. Extraction methods 3a-EDTA LysC Acidic and 5a-EDTA + GuHCl Acidic were also excluded from the study as no reliable MS spectra were generated.

The GdC specimens generated a higher number of MS1 scans and a lower number of MS2 scans than the Ranis specimens by all extraction methods. Hence, Ranis specimens generated more MS2-identified peptides than poorly preserved bone specimens. This difference in MS2-identified scans is because, in DDA mode, GdC specimens do not trigger MS2 scan events as frequently as the minimum ion intensity threshold is only irregularly reached, probably because of highly fragmented peptides based on the preservation of the specimens.

In ideal conditions, deamidation per specimen is expected to show the same value regardless of the extraction method, as the subsamples are derived from a single, homogenized bone powder^{30,51,52}. Nonetheless, this pattern was not shown in our set of specimens. Therefore, our study demonstrates that protein extraction conditions have variable impacts on glutamine and asparagine deamidation. Due to differences in preservation, we observed a variety of deamidation values between the two sets of specimens. However, our estimates of deamidation rates for the GdC specimens cannot be compared properly due to the low number of identified MS2 spectra. The literature on relative glutamine (Q) and asparagine (N) deamidation rates indicates that glutamine residues deamidate at a slower rate than asparagine residues⁵³, as observed in previous palaeoproteomic studies⁵. It is therefore of note that, in our study, we observed that asparagine deamidation ratios for method 6 are below those obtained for glutamine. This pattern, which is the reverse of theoretical expectations, as well as the reverse of most experimentally observed datasets, is difficult to explain. It is, however, consistently present among the Ranis specimens, possibly in the GdC extracts too, and extends to the semi-tryptic search (Supplementary Fig. S10, SI). We therefore believe the observed pattern of more advanced deamidation for glutamines compared to asparagines might be due to the reduction/alkylation step unique to the SPIN protocol in our comparison. Further work in the palaeoproteomics research community should explore this phenomenon in other extraction methods and sample contexts.

In our study, proteomic characteristics were evaluated, such as cleavage rates, NCP genes, and hydrophobicity, to ensure that the extraction methods do not introduce biases into the recovered proteome. The cleavage rate varied between the extraction methods and sample sets in the specific search. The most abundant identified peptides were assigned to collagen. Most of the extraction methods identified no NCPs in any of the specimens. The pattern of NCPs in the Ranis specimens might be explained based on the nature of each protocol. Novel extraction approaches for ZooMS focus on capturing collagenous proteins due to the specificity of the technique, while SPIN was initially designed to retrieve 20 collagenous and non-collagenous protein-coding genes. Additionally, the hydrophobicity of the peptide extracts was calculated as the average of the hydrophobicity of the identified amino acids⁵⁴. Based on the difference in identified peptide intensity, it is shown that different amounts of hydrophilic peptides were extracted by each extraction method. Extraction methods 1, 2, and 3b indicated a bias towards hydrophilic peptides, while 4a, 4b, 5b, and 6 indicated a bias towards hydrophobic peptides. In particular, SPIN showed a significant bias towards hydrophobic peptides. This was probably due to the PAC extraction step, as magnetic sulfur beads tend to capture hydrophobic proteins based on the chemical characteristics of the ligand on the surface of the microbeads.

We presented a comparison of destructive proteomic extraction approaches for species identification by LC-MS/MS. Our results showed that preservation conditions should be taken into consideration when designing proteomic extractions from archaeological bone specimens. Therefore, a pilot study that compares a few extraction approaches based on sample preservation is advised for optimal results. We demonstrated that the acid protein extraction method produces high-quality spectra for bone proteome analysis by LC-MS/MS for degraded samples, would allow for an easy scale-up, and could be preceded by ammonium bicarbonate buffer extraction method, as performed in several ZooMS studies^{7,17,18}. For degraded specimens, we also suggest loading a higher amount (over 10%) of the resulting peptide solution, to enhance MS2 acquisition. An additional step measuring protein concentration in several samples before Evtotip loading might be beneficial to calculate the necessary amount of peptides for optimal LC-MS/MS runs. Moreover, especially in challenging contexts such as the Late Pleistocene, our study showed that the adjustment of data analysis should be considered when designing HTS proteomic studies from archaeological bone specimens. A thorough understanding of the impacts of laboratory processing protocols and analysis methods on the reconstructed proteome is thereby essential to retrieve a maximum amount of unbiased proteomic information from archaeological specimens.

Methods

Specimen selection and sampling

We randomly selected a total of 12 morphologically unidentified bone specimens from two different archaeological sites (Table 1 and Supplementary Fig. S1, SI); six bone specimens from GdC (Portugal) and six bone specimens from Ranis (Germany). Both sets of archaeological specimens date to the Late Pleistocene (LP) and derived from cave sites, with those from GdC expected to be older³ (> 50 kya) and preserved at comparatively high, Mediterranean temperature conditions. GdC is situated on a hill at a height of 20 m above the current sea level, primarily composed of Jurassic and Dolomitic limestones. In contrast, Ranis is slightly younger (unpublished data; 40–50 kya) with comparatively low, central European temperature conditions. The Ranis bone specimens come from the Layer X (Graue Schicht) of the 1932–1938 excavation⁵⁵. This layer is a dark grey humic loamy silt formed in a cave sedimentary environment. The archaeological and paleontological material represents a mix of bone accumulations by carnivores as well as by short-term human occupations. The GdC bone specimens derive from Chamber 2, where the materials used in this study were recovered at the base of a loose sediment talus. The infill is composed of calcareous sediment, limestone blocks collapsed from the cave ceiling, terra rossa, Middle Palaeolithic stone tools, as well as faunal remains⁵⁶. Specimens from GdC displayed variable but generally poor proteome recovery in previous SPIN analysis³, while collagen preservation and ZooMS identification rates at Ranis indicate well-preserved skeletal proteomes (unpublished results). These prior analyses also provided some

taxonomic information for most of the specimen, despite all 12 bone specimens being unidentifiable based on morphological characteristics (Table 1).

Sampling took place in a flow box at the Max Planck Institute for Evolutionary Anthropology (Leipzig, Germany). Surfaces were covered with aluminium foil, which was replaced after sampling each specimen. The flow box, the drilling equipment, and other utensils were cleaned with 70% ethanol and air-dried before use. The samples were drilled into a fine powder using a dental drill. Drill heads were sonicated in 70% ethanol for 10–15 min and air-dried before every use. Approximately 100 mg of bone powder for each sample was generated and homogenized. Subsequently, 5 mg of each sample was placed into a clean 1.5 ml Lo-bind protein Eppendorf tube for each extraction. The remaining bone powder was stored in a separate clean 1.5 ml Lo-bind protein Eppendorf tube at room temperature for potential future analysis.

Proteome extraction methods

We compared the proteome recovery of six proteomic extraction methods (Fig. 1 and Supplementary Table S1, SI). Proteome extraction procedures took place at the Center of Protein Research (Copenhagen, Denmark), where standard laboratory procedures minimized the risk of modern contamination. We assessed two simple methods broadly used for ZooMS, which use ammonium bicarbonate (AmBic or ABC)⁴² and a conventional acid demineralization extraction method⁴. Additionally, two EDTA-based methods with modifications in the digestion enzyme^{5,43,44,57} were tested in this study. We also included a digestion protocol only with LysC. Even though it is commonly used in modern proteomics⁵⁸, studies of ancient proteins demonstrate that these proteins are frequently hydrolyzed to fragment sizes smaller than those encountered in modern proteins⁵⁹. Therefore, the digestion using LysC and trypsin sequentially could, theoretically, lead to an over-cleavage of the surviving protein fragments. Finally, a PAC extraction method³ and an EDTA-based method that combines demineralization and denaturation steps followed by protein digestion *in situ*^{45,46} were also compared.

Here, *in situ* digestion refers to a digestion where both the insoluble bone pellet, left after denaturalization, and the soluble protein solution, after denaturation, are still present. In solution digestion refers to a digestion of proteins in solution after denaturalization only. A brief description of each extraction protocol is given below, whilst all materials and equipment used for this study are shown in Supplementary Table S2, SI. Negative controls were included for each extraction method.

1. Acid-in solution digestion. Samples were demineralized with 100 μ L 5% HCl overnight, at RT. Demineralized samples were centrifuged at 10,000 \times g for 10 min and the supernatant was discarded. The acid-insoluble residue was then washed 3 times with 100 μ L 50 mM ammonium bicarbonate and the pH was ensured to be at 8.0 using pH paper sticks. Then, samples were denatured at 65 °C in 100 μ L 50 mM ammonium bicarbonate for 1-h incubation. Following denaturation, the samples were centrifuged at 10,000 \times g for 10 min to precipitate ungelatinized protein. Fifty μ L of the supernatant was transferred to a clean 96-well plate and the samples were digested with 1 μ L of 50 mM trypsin solution overnight at 37 °C. After digestion, samples were centrifuged at 10,000 \times g for 1 min and acidified using 5 μ L of 5% TFA. The samples were centrifuged at 10,000 \times g for 1 min and the supernatant was transferred to Evtotips for peptide purification and LC–MS/MS analysis.
2. Ammonium bicarbonate- in solution digestion. Five mg of homogenized bone powder was suspended and denatured in 100 μ L 50 mM ammonium bicarbonate pH 8.0 at 65 °C for 1 h incubation. Following denaturation, the samples were centrifuged at 3,000 \times g for 10 min. Fifty μ L of the supernatant was transferred to a clean 96-well plate and digested with 1 μ L of 50 mM trypsin solution overnight at 37°C. After digestion, samples were centrifuged at 10,000 \times g for 10 min and acidified using 5 μ L of 5% TFA. The samples were centrifuged at 3000 \times g for 5 min and the supernatant was transferred to Evtotips for peptide purification and LC–MS/MS analysis.
3. EDTA LysC- *in situ* digestion. The samples were suspended in 100 μ L 0.5 M EDTA pH 8.0 and incubated overnight, at RT with gentle agitation on a shaker. After demineralization, 1 μ L of 50 mM LysC was added to each sample. The samples were digested overnight, at 37 °C with gentle agitation on a thermoshaker. After centrifugation at 10,000 \times g for 10 min, half of the supernatant was acidified with 5 μ L of 5% TFA, and the supernatant was transferred to a new Stagetip (3a, excluded from the study), and the other half was transferred to a Stagetip (3b). After Stagetip clean-up, the peptides were loaded to separate Evtotips for LC–MS/MS analysis.
4. EDTA Protease mix- *in situ* digestion. Samples were suspended in 100 μ L 0.5 M EDTA pH 8.8 and incubated overnight, at RT with gentle agitation on a thermoshaker. After demineralization, the samples were removed from the thermoshaker and 3 μ L of protease mix LysC/Trypsin (1:2) was added to each sample. The samples were digested overnight, at 37 °C with gentle agitation. After centrifugation at 10,000 \times g for 10 min, half of the supernatant was acidified with 5 μ L of 5% TFA and the supernatant was transferred to a new Stagetip (4a), and the other half was transferred to a Stagetip (4b). After Stagetip clean-up, the peptides were loaded to separate Evtotips for LC–MS/MS analysis.
5. EDTA + GuHCl- *in situ* digestion. The samples were demineralized and denatured with 600 μ L solution of 0.5 M EDTA and 3 M GuHCl overnight, at RT. The demineralized samples were centrifuged at 10,000 \times g for 10 min and the supernatant was discarded. The acid-insoluble residue was washed with 100 μ L 50 mM ammonium bicarbonate until the pH was at 8.0. 175 μ L of 50 mM ammonium bicarbonate and 1.5 μ L of protease mix of LysC/Trypsin were added to the sample. After centrifugation at 10,000 \times g for 10 min, half of the supernatant was acidified with 5 μ L of 5% TFA, and the supernatant was transferred to a new Stagetip (5a, excluded from the study), and the other half was transferred to a Stagetip (5b). After Stagetip clean-up, the peptides were loaded to separate Evtotips for LC–MS/MS analysis.

6. SPIN- in solution digestion. The samples were suspended in 100 μL 5% HCl and 0.1% NP-40 solution for overnight demineralization at RT with continuous shaking at 1000 rpm. Reduction, alkylation, and collagen denaturation were facilitated by adding 10 μL 0.1 M tris(2-carboxyethyl) phosphine (TCEP) and 0.2 M N-ethylmaleimide (NEM) in 50% ethanol and 50% ultrapure water and shaking at 1000 rpm at 60 $^{\circ}\text{C}$, for 1 h. The purification and digestion took place on a KingFisherTM Flex (ThermoFisher Scientific) magnetic bead-handling robot. Debris was removed from the protein extract by centrifuging the plate at 800 $\times g$, for 5 min. Magnetic SiMAG-Sulfon beads were washed and prepared at a final concentration of 5 mg/mL in 60% ACN. In a deep-well KingFisherTM plate, 10 μL bead solution and 40 μL of the clear protein extract were briefly mixed. Protein aggregation capture (PAC) was initiated by the addition of 240 μL 70% acetonitrile (ACN; 60% final concentration). The robot was loaded with plates “wash I” (500 μL 70% acetonitrile, 30% water), “wash II” (500 μL 80% ethanol, 20% water), “wash III” (500 μL 100% acetonitrile, and the “on-bead-digestion” plate (100 μL 20 mM Tris pH 8.5, 1 $\mu\text{g}/\text{mL}$ LysC, 2 $\mu\text{g}/\text{mL}$ Trypsin). The programmed sequence was: (i) collect the beads with low speed for 3:30 min, (ii–iv) wash I–III with slow mixing for 2 min, and (v) bead release on the digestion plate. The digestion was finalized outside the robot, shaking at 800 rpm at 37 $^{\circ}\text{C}$, overnight. The peptides were acidified with 10 μL 5% trifluoroacetic acid (TFA). After acidification, the peptides were purified in an Evotip for LC–MS/MS analysis.

Peptide purification and Evotip loading

The purification of the digested peptides was performed for all non-EDTA-based extraction methods directly purified on Evotips as described below (methods 1, 2, and 6). However, EDTA in acidic conditions (methods 3, 4, and 5) usually results in precipitates and might block the Evotips, negatively affecting downstream peptide elution and data acquisition during mass spectrometry. To minimize the chances of this happening, all EDTA peptide digestions were split in half and eluted on homemade Stagetips as follows, before Evotip loading. The other half was purified directly to the Evotips. The extraction approaches 3a and 5a generated no MS data and they are excluded from this study.

- a. *Stagetip equilibration and loading*: One Stagetip was activated per sample. All Stagetips were activated with 50 μL 100% ACN and equilibrated twice with 50 μL 5 mM AmBic. The supernatant was removed between all the steps with centrifugation at 700 $\times g$ for 60 sec. The equilibrated tips were loaded with the supernatant and washed twice with 50 μL 5 mM AmBic, the pH was adjusted by washing twice with 50 μL of 0.1% TFA and the peptides were eluted with 50 μL 40 % ACN and 50 μL 60 % ACN. Finally, ACN was removed from the eluate with a speed vac.
- b. *Evotip equilibration and loading*: One Evotip (Evosep, EV-2001) per sample was washed in ACN, soaked with isopropyl alcohol, and equilibrated with 0.1% FA in water, according to the manufacturer’s protocol. The equilibrated tips were loaded with 10% peptide solution (out of 100 μL) and washed with 20 μL 0.1% FA before LC-MS/MS.

LC–MS/MS for palaeoproteomics

LC–MS/MS was carried out using the 60 samples per day (SPD) DDA method of an Evosep One (Evosep, Odense, Denmark⁴¹) operated with the Evosep plugin (1.4.381.0) in Chronos (4.9.2.0) and an analytical column made in-house using a laser-pulled 16 cm long 150 μm inner diameter capillary packed with 1.9 μm C18 bounded silica particles (ReproSil-Pur, C18-AQ, Dr. Maisch, Germany). The column was mounted on an electrospray source with a column oven set at 60 $^{\circ}\text{C}$. Peptides were ionized by nano-electrospray at 2 kV and analyzed on an Orbitrap Exploris 480[™] (Thermo Fisher Scientific, Bremen, Germany) MS operated with Xcalibur (3.1–4). Full scans ranging from 350 to 1400 m/z were measured at 60 k resolution, 25 ms max. IT, 300% AGC target. The top 8 precursors were selected (30 s dynamic exclusion) for HCD fragmentation with an isolation window of 1.3 m/z and a NCE of 30. The minimum intensity to trigger an MS2 Scan was lowered to 5e4. MS2 scans were acquired at 15 k resolution, 22 ms max. IT, and 200% AGC target. Using such a short gradient DDA approach reduces the cost per sample by an order of magnitude compared to previous LC–MS/MS-based species identification strategies. In addition to higher throughput and lowered analysis costs, short gradients in the EvoSep One allowed for stable storage of the samples on EvoTips, avoiding preparation steps like peptide elution after cleanup, solvent evaporation, and MALDI target plate spotting.

MaxQuant search

All .raw files were analyzed in MaxQuant version 2.1.1.0 against a reference protein sequence database provided by Ruther et al. 2022³. Download dates are available in the MaxQuant “summary.txt” files for ‘specific’ and ‘semi-specific’ searches (PXD042321- ProteomeXchange online repository). Variable modifications were included, such as oxidation (M), deamidation (NQ), Gln- > pyro-Glu, Glu- > pyro-Glu, and proline hydroxylation, whereas NEM-derivatization of Cys was configured as a fixed modification for method 6. Initial searches were run in specific Trypsin/P digestion mode and allowed for up to two miscleavages. All files were searched against the above-mentioned database using the same settings. Up to five variable modifications were allowed. The internal MaxQuant contaminant list was replaced with a custom database³. All other settings were left on default. A second search was performed using semi-specific Trypsin/P digestion mode and up to four miscleavages were allowed, with all other settings left identical.

Data analysis

After spectral identification, data analysis was conducted largely through R (version 4.1.2) in RStudio (version 2022.02.0.0) using the packages tidyverse (version 1.3.1)⁶⁰, seqinr (version 4.2–8), devtools (version 2.4.4), ggpubr (version 0.4.0), data.table (version 1.14.2)⁶¹, bit64 (version 4.0.5)⁶², ggsci (version 2.9), progressr (version 0.10.0)⁶³, gmp (version 0.6–6)⁶⁴, reshape2 (version 1.4.4)⁶⁵ and stringi (version 1.7.6). Deamidation was quantified based on spectral intensities, following Mackie et al.⁵¹. Hydrophobicity values (a measure of hydrophobicity) were estimated for recovered proteins using the web application GRAVY Calculator (www.gravy-calculator.de/) and Peptides (version 2.2.4)⁶⁶. Statistics were calculated using the two-way ANOVA (Type II) tests⁶⁷ from carData (version 3.0–5) and car (version 3.1–0)⁶⁸. The map was built with the package maps (version 3.4.1)⁶⁹.

Data availability

The raw mass spectrometry proteomics data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE⁷⁰ partner repository with the dataset identifier PXD042321.

Received: 22 May 2023; Accepted: 13 October 2023

Published online: 26 October 2023

References

- Sharif, M. B. et al. Ancient DNA refines taxonomic classification of Roman equids north of the Alps, elaborated with osteomorphology and geometric morphometrics. *J. Archaeol. Sci.* **143**, 105624 (2022).
- Dai, Y. et al. Identification of cetaceans from bones using molecular techniques provides insights into cetacean species diversity and composition in coastal western Taiwan Strait waters, China. *Conserv. Genet. Resour.* <https://doi.org/10.1007/s12686-022-01296-x> (2022).
- Rüther, P. L. et al. SPIN enables high throughput species identification of archaeological bone by proteomics. *Nat. Commun.* **13**, 1–14 (2022).
- Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854. <https://doi.org/10.1002/rcm.4316> (2009).
- Cappellini, E. et al. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *J. Proteome Res.* **11**, 917–926. <https://doi.org/10.1021/pr200721u> (2012).
- Sawafuji, R. et al. Proteomic profiling of archaeological human bone. *R. Soc. Open Sci.* **4**, 161004 (2017).
- Cappellini, E. et al. Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantiidae). *Zool. J. Linn. Soc.* **170**, 222–232 (2014).
- Hublin, J.-J. et al. Initial upper palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. *Nature* **581**, 299–302 (2020).
- Brown, S. et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6**, 23559 (2016).
- Sinet-Mathiot, V. et al. Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy). *Sci. Rep.* **9**, 12350 (2019).
- Welker, F. et al. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11162–11167 (2016).
- Welker, F., Soressi, M., Rendu, W., Hublin, J.-J. & Collins, M. Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *J. Archaeol. Sci.* **54**, 279–286 (2015).
- Ruebens, K. et al. The Late Middle Palaeolithic Occupation of Abri du Maras (Layer 1, Neronian, Southeast France): Integrating lithic analyses, ZooMS and radiocarbon dating to reconstruct neanderthal hunting behaviour. *J. Paleolit. Archaeol.* **5**, 4 (2022).
- Culley, C. et al. Iron Age hunting and herding in coastal eastern Africa: ZooMS identification of domesticates and wild bovids at Panga ya Saidi. *Kenya. J. Archaeol. Sci.* **130**, 105368 (2021).
- Ruebens, K. et al. Neanderthal subsistence, taphonomy and chronology at Salzgitter-Lebenstedt (Germany): A multifaceted analysis of morphologically unidentifiable bone. *J. Quat. Sci.* <https://doi.org/10.1002/jqs.3499> (2023).
- Pothier Bouchard, G., Riel-Salvatore, J., Negrino, F. & Buckley, M. Archaeozoological, taphonomic and ZooMS insights into The Protoaurignacian faunal record from Riparo Bombrini. *Quat. Int.* **551**, 243–263 (2020).
- Bradfield, J., Kitchener, A. C. & Buckley, M. Selection preferences for animal species used in bone-tool-manufacturing strategies in KwaZulu-Natal, South Africa. *PLoS One* **16**, e0249296 (2021).
- McGrath, K. et al. Identifying archaeological bone via non-destructive ZooMS and the materiality of symbolic expression: Examples from Iroquoian bone points. *Sci. Rep.* **9**, 11027 (2019).
- Martisius, N. L. et al. Initial Upper Paleolithic bone technology and personal ornaments at Bacho Kiro Cave (Bulgaria). *J. Hum. Evol.* **167**, 103198 (2022).
- Martisius, N. L. et al. Non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals. *Sci. Rep.* **10**, 7746 (2020).
- Talamo, S. et al. A 41,500 year-old decorated ivory pendant from Stajnia Cave (Poland). *Sci. Rep.* **11**, 1–11 (2021).
- Tyanova, S. et al. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453–1456 (2015).
- Bekker-Jensen, D. B. et al. A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Mol. Cell. Proteom.* **19**, 716–729 (2020).
- Ping, P., Hermjakob, H., Polson, J. S., Benos, P. V. & Wang, W. Biomedical informatics on the cloud: A treasure hunt for advancing cardiovascular medicine. *Circ. Res.* **122**, 1290–1301 (2018).
- Bekker-Jensen, D. B. et al. An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst.* **4**, 587–599.e4 (2017).
- Welker, F. et al. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature* **522**, 81–84 (2015).
- Buckley, M. et al. Correction: Collagen sequence analysis of the extinct giant ground sloths *Lestodon* and *Megatherium*. *PLOS ONE* **10**, e0144793. <https://doi.org/10.1371/journal.pone.0144793> (2015).
- Cleland, T. P., Schroeter, E. R., Feranec, R. S. & Vashishth, D. Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proc. Biol. Sci.* **283**, 20160593 (2016).
- Buckley, M. et al. Collagen sequence analysis reveals evolutionary history of extinct West Indies nesophontes (Island-Shrews). *Mol. Biol. Evol.* **37**, 2931–2943 (2020).
- Hendy, J. Ancient protein analysis in archaeology. *Sci. Adv.* <https://doi.org/10.1126/sciadv.abb9314> (2021).
- Harvey, V. L. et al. Preserved collagen reveals species identity in archaeological marine turtle bones from Caribbean and Florida sites. *R. Soc. Open Sci.* **6**, 191137 (2019).

32. Welker, F., Smith, G. M., Hutson, J. M. & Kindler, L. Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene. *PeerJ* <https://doi.org/10.7717/peerj.3033> (2017).
33. Cappellini, E. *et al.* Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature* **574**, 103–107. <https://doi.org/10.1038/s41586-019-1555-y> (2019).
34. Welker, F. *et al.* The dental proteome of *Homo* antecessor. *Nature* **580**, 235–238 (2020).
35. Demarchi, B. *et al.* Ancient proteins resolve controversy over the identity of *Genyornis* eggshell. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2109326119. <https://doi.org/10.1073/pnas.2109326119> (2022).
36. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteom.* **11**, O111.016717 (2012).
37. Geromanos, S. J., Hughes, C., Ciavarini, S., Vissers, J. P. C. & Langridge, J. I. Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. *Anal. Bioanal. Chem.* **404**, 1127–1139 (2012).
38. Geiger, T., Cox, J. & Mann, M. Proteomics on an orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol. Cell. Proteom.* **9**, 2252–2261. <https://doi.org/10.1074/mcp.m110.001537> (2010).
39. Bath, T. S. *et al.* Protein aggregation capture on microparticles enables multipurpose proteomics sample preparation. *Mol. Cell. Proteom.* **18**, 1027–1035 (2019).
40. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).
41. Bache, N. *et al.* A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol. Cell. Proteomics* **17**, 2284–2296 (2018).
42. van Doorn, N. L., Hollund, H. & Collins, M. J. A novel and non-destructive approach for ZooMS analysis: Ammonium bicarbonate buffer extraction. *Archaeol. Anthropol. Sci.* **3**, 281–289. <https://doi.org/10.1007/s12520-011-0067-y> (2011).
43. Cleland, T. P. Solid digestion of demineralized bone as a method to access potentially insoluble proteins and post-translational modifications. *J. Proteome Res.* **17**, 536–542 (2018).
44. Jensen, T. Z. T. *et al.* Tryps-In: A Streamlined Palaeoproteomics Workflow Enables ZooMS Analysis of 10,000-Year-Old Petrous Bones from Jordan Rift-Valley. *J. Arch. Sci.: Reports* **52**, 104238. <https://doi.org/10.1016/j.jasrep.2023.104238> (2023).
45. Schweitzer, M. H. *et al.* Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. canadensis*. *Science* **324**, 626–631 (2009).
46. Schroeter, E. R., DeHart, C. J., Schweitzer, M. H., Thomas, P. M. & Kelleher, N. L. Bone protein ‘extractomics’: Comparing the efficiency of bone protein extractions of *Gallus gallus* in tandem mass spectrometry, with an eye towards paleoproteomics. *PeerJ* **4**, e2603 (2016).
47. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
48. Ramsøe, A. *et al.* DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *J. Archaeol. Sci.* **115**, 105080 (2020).
49. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0) (1982).
50. Le Meillour, L. *et al.* Identification of degraded bone and tooth splinters from arid environments using palaeoproteomics. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **511**, 472–482 (2018).
51. Mackie, M. *et al.* Palaeoproteomic profiling of conservation layers on a 14th century Italian wall painting. *Angew. Chem. Int. Ed Engl.* **57**, 7369–7374 (2018).
52. Ramsøe, A. *et al.* Assessing the degradation of ancient milk proteins through site-specific deamidation patterns. *Sci. Rep.* **11**, 7795 (2021).
53. Robinson, N. E. *et al.* Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J. Pept. Res.* **63**, 426–436 (2004).
54. Eriksson, K. O. Chapter 19—Hydrophobic Interaction Chromatography. In *Biopharmaceutical Processing* (eds Jagschies, G. *et al.*) 401–408 (Elsevier, 2018). <https://doi.org/10.1016/B978-0-08-100623-8.00019-0>.
55. Hülle, W. *Die Ilsenhöhle unter Burg Ranis, Thüringen: einpaläolithische Jägerstation* (Fischer, 1977).
56. Barbieri, A., Regala, F. T., Cascalheira, J. & Bicho, N. The sediment at the end of the tunnel: Geophysical research to locate the Pleistocene entrance of Gruta da Companheira (Algarve, Southern Portugal). *Archaeol. Prospect.* **30**, 117–134 (2023).
57. Orlando, L. *et al.* Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
58. Dau, T., Bartolomucci, G. & Rappsilber, J. Proteomics using protease alternatives to trypsin benefits from sequential digestion with trypsin. *Anal. Chem.* **92**, 9523–9527 (2020).
59. Collins, M. J. *et al.* The survival of organic matter in bone: A review. *Archaeometry* **44**, 383–394 (2002).
60. Wickham, H. *et al.* Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
61. *data.table: R's data.table package extends data.frame*: (Github).
62. *Jens.Bit64: An R package with an S3 Class for Vectors of 64bit Integers*. (Github).
63. Bengtsson, H. *progressr: R package: An Inclusive, Unifying API for Progress Updates*. (Github).
64. Sylvain Jasson / gmp · GitLab <https://forgemia.inra.fr/sylvain.jasson/gmp>.
65. Wickham, H. Reshaping data with the reshape package. *J. Stat. Softw.* **21**, 1–20 (2007).
66. Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: A package for data mining of antimicrobial peptides. *Small* **12**, 44–444 (2015).
67. Box, G. E. P., Hunter, W. G. & Hunter, J. S. *Statistics for Experimenters* (Wiley, 1978).
68. R-Forge: car: Companion to Applied Regression: Project Home. <https://r-forge.r-project.org/projects/car/>.
69. Maps: Draw geographical maps. *Comprehensive R Archive Network (CRAN)* <https://cran.r-project.org/web/packages/maps/index.html>.
70. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research, the innovation programme under the Marie Skłodowska-Curie (grant agreement No. 861389- PUSHH), and the Max Planck Society. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 948365). G.M.S. received funding from the European Union's Horizon Europe Research and Innovation Programme under Marie Skłodowska-Curie Grant Agreement 101027850. A.J. Taurozzi was supported by the Danish National Research Foundation award PROTEIOS (DNRF128). R.M. Godinho is funded by Fundação para a Ciência e a Tecnologia (FCT; contract reference 2020.00499.CEECIND and FCT R&D project reference 2022.07737.PTDC). Bones from the site of Companheira, Portugal, were retrieved during excavations funded by the Portuguese Ministry for Science and Technology (ALG-01-0145-FEDER-27833) and Fundos FEDER through Programa Operacional Regional do

Algarve- CRESA Algarve2020 (PTDC/HAR-ARQ/27833/2017). J Cascalheira is funded by Fundação para a Ciência e a Tecnologia, contract reference DL57/2016/CP1361/CT0026. Current archaeological work at Gruta da Companheira is funded by the European Union (ERC-CoG-FINISTERRA-101045506)—views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. We thank the State Museum of Prehistory Halle (Saale) for providing access to the specimen collection. Thanks to P.L. Rütger, G. Troché, V. Sinet-Mathiot, K. Ruebens and H. Xia for their assistance. We also thank the reviewers and the editor for their help and comments.

Author contributions

D.M., F.S.H., and F.W. designed the research. A.J.T. provided access to one extraction protocol and advised on the methodological comparison. R.M.G., M.W., T.S., S.P.M., G.M.S., J.C., and N.B., contributed archaeological samples. J.-J.H., J.V.O., and F.W. ensured access to the required equipment. D.M. and F.S.H. performed laboratory research. D.M. and F.S.H. analyzed the data. Z.F. and F.W. provided computational interpretations. D.M. and F.W. prepared the manuscript. All co-authors contributed to the revision of the manuscript.

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/s41598-023-44885-y>.

Correspondence and requests for materials should be addressed to D.M. or F.W.

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) 2023