

# Open science, communication, and collaboration for the future of palaeoproteomics

## *Authors:*

Joannes A. A. Dekker<sup>1,2,3</sup>, Carli Peters<sup>4,5</sup>, Rachel M. Winter<sup>6</sup>, Matthew J. Collins<sup>1,7</sup>, Marc R. Dickinson<sup>8</sup>, Virginia L. Harvey<sup>9</sup>, Eden Hill<sup>10,11</sup>, Bharath Nair<sup>7</sup>, Takumi Tsutaya<sup>12</sup>, Laura C. Viñas-Carón<sup>1</sup>, Christina Warinner<sup>13,14,15</sup>, Frido Welker<sup>1</sup>, Zandra Fagernäs<sup>1</sup>

## *Affiliations:*

1. Globe Institute, University of Copenhagen, Copenhagen, Denmark
2. BioArCh, University of York, York, United Kingdom
3. Department of Life Sciences and Systems Biology, University of Turin, Italy
4. Interdisciplinary Center for Archaeology and the Evolution of Human Behaviour, Universidade do Algarve, Faro, Portugal
5. Department of Archaeology, Max Planck Institute of Geoanthropology, Jena, Germany
6. Center for Research on Biodiversity and Environment, University of Toulouse, Toulouse, France
7. McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, United Kingdom
8. Department of Chemistry, University of York, York, United Kingdom
9. School of Natural Sciences, University of Chester, Chester, United Kingdom
10. Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
11. Department of Geosciences, Eberhard Karl University of Tübingen, Tübingen, Germany
12. Research Center for Integrative Evolutionary Science, The Graduate University for Advanced Studies (SOKENDAI), Kanagawa, Japan
13. Department of Anthropology, Harvard University, Cambridge, USA 02138
14. Department of Human Evolutionary Biology, Harvard University, Cambridge, USA 02138
15. Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany 04103

## *Corresponding author:*

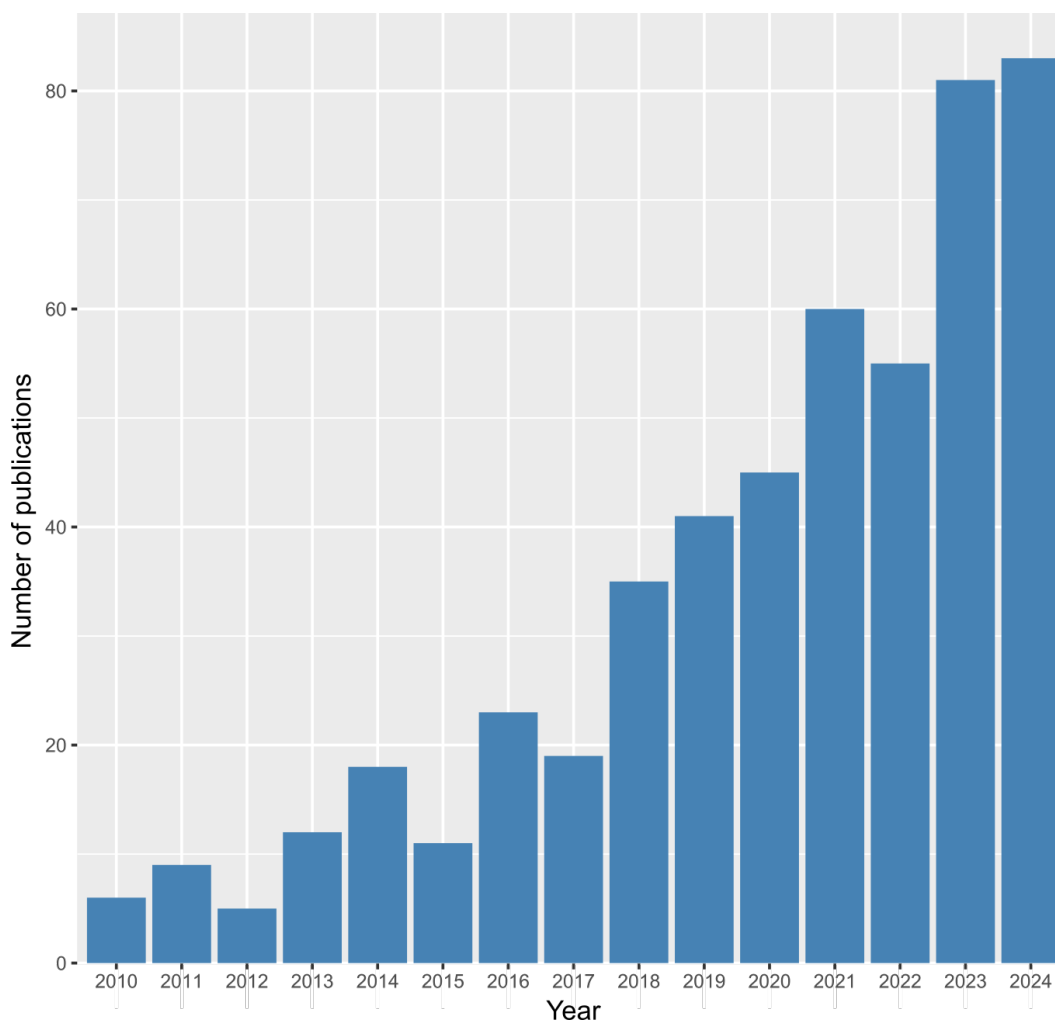
Zandra Fagernäs  
Øster Farimagsgade 5  
Building 7, 1st floor  
1353, Copenhagen  
Denmark  
zandra.fagernas@sund.ku.dk

## Abstract

Palaeoproteomics focuses on studying the past by analysing proteins preserved in historical, archaeological, and palaeontological materials. Over the past fifteen years, the number of publications and research questions in this field has grown substantially, as has the number of researchers and research groups utilising palaeoproteomic methods. PAASTA (Palaeoproteomics And Archaeology, Society for Techniques and Advances) is an international early career researcher-driven community, which aims to bring together researchers from all subdisciplines of palaeoproteomics by providing a platform for open communication. Here, we discuss ways in which this fast-growing research field can benefit from adopting principles of open science and increased communication and collaboration between researchers. We approach this question from all stages of a typical research project, from experimental design and sample selection, to data analysis and raw data sharing. Although we acknowledge that each of these guidelines and recommendations may not be applicable to every research project, we aim to provide a foundation which, when applied collectively, will strengthen the future of palaeoproteomics.

## Introduction

The field of palaeoproteomics concerns the study of the past through proteins preserved in historical, archaeological, and palaeontological materials (Warinner *et al.* 2022). The number of publications per year in this field has increased significantly over the last fifteen years (Fig. 1), with palaeoproteomics being applied to an ever-expanding number of material types and research questions. Broadly, the field can be classified into three categories of questions, each progressively focusing on finer details of the proteome: I) identification of proteins, II) identification of protein sequence variation, and III) identification of amino acid modifications.



**Figure 1.** Number of palaeoproteomics articles published each year since 2010 (excluding preprints). The list of papers was compiled by a Web of Science query (accessed on 14 June 2024) using the search terms: “palaeoproteomics” OR “paleoproteomics” OR “ancient proteins” OR “ZooMS” OR “collagen fingerprinting” OR “peptide mass fingerprinting”. Preprints, meeting abstracts, and entries that did not relate to the study of ancient proteins were removed from the WoS dataset. Palaeoproteomic papers that were missing from this list and papers that were published after the WoS search date were manually added to the dataset. This figure was made in R v.4.4.2. with R Studio v.2024.12.0+467 (R Core Team 2023) using ggplot2 (Wickham 2016). See also Supplementary Dataset 1 for a full record of the curated dataset.

A large fraction of palaeoproteomic studies utilise the detection of a specific protein to identify certain events or practices from the past. For example, the detection of proteins relating to milk consumption from dental calculus is a way of identifying the use of animals for dairying in the past (Hendy *et al.* 2018a; Scott *et al.* 2022; Warinner *et al.* 2014a). Similarly, proteins deriving from pathogens or the immune response can be used to identify the presence of disease (Fotakis *et al.* 2020; Warinner *et al.* 2014b; Wilkin *et al.* 2024), while the presence of one or both of the amelogenin proteins, which are encoded on the sex chromosomes, can be used to assign genetic sex (Stewart *et al.* 2017). On a wider scale, metaproteomics utilises all proteins present in a sample to collectively study the material (Hendy *et al.* 2018a; Jersie-Christensen *et al.* 2018). Specific questions about the past can thereby be answered through the identification of proteins in different materials.

Amino acid sequence variation in proteins between species has led to palaeoproteomic approaches for taxonomic identification being developed, such as ZooMS (Zooarchaeology by Mass Spectrometry; Buckley *et al.* 2009) and SPIN (Species by Proteome INvestigation; R  ther *et al.* 2022). These allow for the taxonomic identification of skeletal or other faunal remains (such as parchment, textiles or artefacts) even when no morphologically identifiable characteristics are present, or when osteomorphology is too similar between taxa for distinctions. Furthermore, variation in amino acid sequences between taxa also allows for phylogenetic placement (Buckley 2015; Cappellini *et al.* 2019; Demarchi *et al.* 2022; Harvey *et al.* 2021; Lowenstein *et al.* 1981; Welker *et al.* 2015; 2020), enabling reconstruction of phylogenetic trees and hypotheses about evolutionary relationships between extinct and extant species. Long-standing questions in other research fields, such as zooarchaeology, palaeoanthropology, historical ecology, or zoology, can thereby be addressed through the study of protein sequence variation.

Finally, identification of post-translational modifications (PTMs) of amino acids can provide valuable information about the history and preservation of the proteins. For example, deamidation of asparagine and glutamine may be used to assess the relative antiquity of a detected protein or peptide (Brown *et al.* 2021b; Rams  e *et al.* 2020; Wilson *et al.* 2012), and for more specific purposes, such as investigating manufacturing differences in parchment (Nair *et al.* 2023). The extent of photo-oxidative modifications can lead to inferences about past light exposure of artistic or cultural heritage materials, such as paintings (Mackie *et al.* 2018). Additionally, amino acid racemisation—the spontaneous interconversion of L-amino acids to their D-enantiomers over time—provides a valuable chronological tool for age determination. Amino acid geochronology has been successfully applied to various materials, such as shells, eggshells, coral, and enamel, enabling relative dating in palaeontological and archaeological contexts (Dickinson *et al.* 2024; Miller *et al.* 1992; Penkman *et al.* 2011; Wehmiller *et al.* 1976). Evaluating the extent of protein decomposition, through processes like racemisation, can be useful for screening potential samples for proteomic analysis (Presslee *et al.* 2021), and confirming the presence of endogenous amino acids (Cappellini *et al.* 2019), a key consideration for older or more degraded samples at the limits of palaeoproteomic analysis. More broadly, the survival of stable amino acids and the presence of degradation products, first explored by Abelson (1954), is evidence for the persistence of protein remnants into deep time (Saitta *et al.* 2024).

The field of palaeoproteomics is thereby highly diverse in terms of both research questions and methodological approaches. However, researchers in this field commonly face several issues. For example, ancient proteomes are inherently complex and often contaminated. Even in cases where only one protein or peptide is the object of research, proteins stemming from the burial environment and subsequent storage and handling may be present on the studied material, or introduced during the extraction process (Fagernäs *et al.* 2024a; Hendy *et al.* 2018b). Ancient proteins are also degraded and damaged, complicating analyses (Hendy 2021). Finally, the materials that are studied through palaeoproteomics are often rare and/or of significant cultural heritage value. Precautions should be taken and efforts made to minimise the damage that is caused when sampling for palaeoproteomic analyses. Since these issues, and many more, are shared by this research community, the solutions can also be shared.

PAASTA, short for *Palaeoproteomics And Archaeology, Society for Techniques and Advances*, is an international, early career researcher-driven community of palaeoproteomics researchers. Founded in 2023 as an affiliate of the *International Society for Biomolecular Archaeology* (ISBA), PAASTA follows the example of other successful early career researcher-driven communities like SPAAM (*Standards, Precautions, and Advances in Ancient Metagenomics*). PAASTA's main mission is to promote greater collaboration and transparency within the field. By fostering a sense of community, we hope to minimise the risk of encountering some of the challenges that have emerged, especially for early career researchers, in the field of ancient DNA, through competition and lack of communication (Callaway 2017; Horsburgh 2024; Jones 2022; Källén *et al.* 2024). By providing a platform for researchers with an interest in palaeoproteomics to communicate with each other in a supportive manner, the community aims to improve connections between groups at different research institutes and across the world. In this manner, we aim to advance palaeoproteomics by sharing knowledge and jointly finding solutions to problems faced by many researchers in the field. In practice, this is achieved by providing an online space for communication, organising events such as virtual seminars and an annual conference, as well as facilitating collaborative projects within the community. All researchers are welcome to join the community, irrespective of research focus or career stage, but we particularly encourage an open space for early career researchers to ask questions and engage in discussions.

At the core of PAASTA is an aim to make the emerging and fast-developing field of palaeoproteomics more open, communicative, and collaborative. Here, we walk through the usual steps of a palaeoproteomics project, and discuss how each step could align with these principles, while also integrating the FAIR (guidelines for making data *Findable, Accessible, Interoperable and Reusable*; Wilkinson *et al.* 2016) and CARE (guidelines for sharing Indigenous Peoples data; *Collective benefit, Authority to control, Responsibility and Ethics*; Research Data Alliance International Indigenous Data Sovereignty Interest Group 2019) principles. We are aware that each archaeological science project is unique and all suggestions below are unlikely to be applicable to all projects. However, by discussing and promoting different options and guidelines, we can establish a baseline for this community and collectively strengthen the research field step-by-step.

## Experimental Design

The first step in any palaeoproteomics project is setting up an appropriate research design and sampling strategy that can answer the specific research question asked. Such a strategy should be thoroughly discussed with all relevant stakeholders and planned prior to sampling. Various factors should be taken into account, such as preservation differences within samples (Ásmundsdóttir *et al.* 2024) or sites (Le Meillour *et al.* 2024), sample treatment (e.g., legacy samples or museum specimens) and manufacturing processes (e.g. liming, tanning, dyeing). Curators have become increasingly wary about destructive analysis as a result of the “... *tension between the professional obligations of curators to faithfully steward archaeological collections and ensure their ongoing integrity and the “harvesting” desires of specialists intent on industrial-scale aDNA research in the absence of a research question*” (Källén *et al.* 2024). To avoid a breakdown in relationships between communities, archaeologists and museums, we must ensure that destructive sampling is minimised and additional research opportunities have been considered. Given that the methods for extracting microfossils, ancient DNA (aDNA), proteins and lipids target the different physicochemical characteristics of each material, it should be the intention to sample in a way that maximizes the potential to generate complementary data from the same sub-samples (e.g., for archaeogenetic analysis, isotopic analysis, radiocarbon dating).

Palaeoproteomic analyses are often destructive to a certain degree, as part of a specimen is irreversibly removed, but the degree of destructiveness varies depending on the chosen extraction protocol. Non-destructive protocols have also been developed, but vary in their success levels (Hansen *et al.* 2024). Open and collaborative communication with those that provide us with material to sample (communities, archaeologists, conservators, collection managers, etc.), addressing the risks of the study and the extent of sampling destructiveness (e.g., the amount of material required and what will remain), is essential for balancing the needs of palaeoproteomics research with the conservation of objects or remains (Fleskes *et al.* 2022; Pálsdóttir *et al.* 2019). During the sampling process, it is also recommended to take high-quality photos before and after sampling of the specimen; additionally, institutions curating human remains may require 3D or CT scanning of remains prior to destructive analysis, which adds to the cost. Documentation prior to and after sampling will both aid in identifying visual indicators that distinguish successful/unsuccessful samples and promote transparency regarding the traces left by sampling methods on ancient specimens.

Collaboration across different research specialties is both essential and highly beneficial for the success of a research project. When working with material that originates directly from an archaeological site, it is vital to work collaboratively with the site specialists, excavation teams, and any other specialists involved. They have the site-specific knowledge about stratigraphy, taphonomy, and chronology of the archaeological contexts, all of which need to be taken into account when selecting which material is suitable to sample for palaeoproteomics research. The value of any archaeological object for our understanding of the past is highly dependent on the archaeological context it was found in. Engaging with archaeologists, curators or other stakeholders from the earliest stages of experiment design can be immensely helpful in ensuring that the selected samples match the research questions and that the research questions

themselves will meaningfully contribute to advancing our understanding of the past. For example, in the case of ZooMS and SPIN, active collaboration with zooarchaeologists that work with the faunal material is crucial to the development of a suitable sampling strategy that not only takes into account the palaeoproteomic perspective, but that also helps address specific research questions stemming from zooarchaeology (Hansen *et al.* 2024). Additionally, an understanding of the faunal spectrum is essential for maximising the specificity of the taxonomic identifications made with methods like ZooMS and SPIN. Whilst certain taxa may share the same peptide sequences, they can often be distinguished based on their temporal and geographic distribution. Further, closer collaboration with specialists in other fields in biomolecular archaeology (e.g., isotope ratio mass spectrometry, archaeogenomics, and organic residue analysis) can lead to the development of simultaneous or sequential extraction protocols that minimise the need for duplicated destructive sampling (Fagernäs *et al.* 2020; Rusu *et al.* 2019). Finally, collaboration with bioinformaticians and statisticians can help enhance data analysis and interpretation, as well as ensure that the research design has the power to answer the research question.

The rights of Indigenous peoples and the colonial nature of resource-intensive ancient biomolecular research have been much debated in the field of archaeogenetics (Fleskes *et al.* 2022; Källén 2025; Källén *et al.* 2024; Somel *et al.* 2021). As an emerging field, palaeoproteomics can learn from the discussion on the ethics of aDNA research and has the potential to democratise ancient biomolecular research by promoting communication and collaboration with stakeholders and researchers in the country of origin of the samples (Paterson *et al.* 2024). Participation in palaeoproteomic research can also be broadened through citizen science projects (Brandt *et al.* 2022). Since proteins can be analysed from materials that are significantly older than those used for ancient DNA studies (Cappellini *et al.* 2019; Demarchi *et al.* 2022; Harvey *et al.* 2018; Welker *et al.* 2020), the ethical dimensions of palaeontological research, which have not been much discussed in ancient DNA research, may become a challenge for palaeoproteomics. For example, palaeontological specimens are sometimes commercially traded or privately owned, which is more rarely the case in archaeology. The research use of such specimens is debated in palaeontology (Haug *et al.* 2020; Rayfield *et al.* 2020; Santucci *et al.* 2016), and an open interdisciplinary discussion among researchers within palaeoproteomics is essential to establish clear ethical guidelines that balance scientific advancement with respect for the cultural and legal complexities surrounding these materials.

## Protein Extraction and Data Acquisition

Protein extraction protocols are continuously being developed to increase the amount of high-quality data that is obtained from each sample. Sharing laboratory protocols is essential for reproducibility of studies and for advancing the field as a whole. Methods sections in traditional journal publications are, however, often limited in the amount of detail that can be included. Recreating a protocol based on methods reported within an article can therefore be highly challenging. Furthermore, citations to methods in previous publications often lead to issues with finding the original protocol, including broken citation chains and publications being paywalled (Standvoss *et al.* 2024). Sharing detailed stepwise protocols on open platforms, such as

protocols.io, ensures that the method can be accurately adapted by other researchers, while also ensuring that the original publication is cited, and that different adaptations of the same protocol can be traced using a version-tracking system (e.g. Brown *et al.* 2020; Poujois & Le Meillour 2023; Scott & Warinner 2020). For methods sections in journal articles, guidelines have been created for what should be included (Giraldo *et al.* 2018), containing details ranging from the protocol version that is used, to the precise equipment, consumables and reagents, and exact procedures used at each step. This level of detailed method sharing ensures that for each destructive sampling of archaeological specimens, the most appropriate protocol can be selected and applied, irrespective of which research group is analysing the specimen.

Considering the unique nature of archaeological samples, it can be difficult to evaluate the effectiveness of a protocol at extracting proteins, even if its details have been shared fully. A more widespread inclusion of positive controls or standards would facilitate the comparison of protein extraction between different protocols. These might include standards added to every MALDI plate to overcome issues related to batch effects, as well as the inclusion of biological or technical duplicates to assess the efficiency of each protein extraction and data acquisition batch (Viñas-Caron *et al.* 2023). Ideally, the same protein standard or positive control would be shared between different research groups in order to also enable inter-laboratory comparisons. The PAASTA community can serve as a forum to facilitate discussions on selecting appropriate materials for positive controls, ensuring that decisions are collectively made and that the protein standards are accessible to all members of the community. In a similar manner, the robustness of results can be enhanced by replication in different laboratories, by different researchers, which the PAASTA community is ideally placed to facilitate.

Mass spectrometry methods are also constantly developing and improving; however, researchers in the field of palaeoproteomics are rarely involved in such methodological improvements to instrumentation, in part because many palaeoproteomics groups rely on core facilities where they may have limited input or control over instrument operations. Closer collaboration with researchers in the broader fields of protein mass spectrometry and proteomics may lead to significant steps forward in palaeoproteomics due to the faster adoption of new methodologies. Moreover, such collaborations might also lead to the development of methods specifically designed to suit the challenges of palaeoproteomics. For example, SPIN (Rüther *et al.* 2022) was developed for palaeoproteomics in parallel to applications such as bead-based protein capture and digestion, EvoSep analysis (Bache *et al.* 2018), and DIA data acquisition (Venable *et al.* 2004) being developed for clinical proteomics. The development of EvoSep liquid chromatography column systems also facilitates the standardisation of the chromatographic separation of proteins prior to MS/MS data acquisition. The limited number of standardized EvoSep methods available for selection in EvoSep makes it easy to replicate precise settings. Although this does reduce the amount of optimization that can be performed by changing settings, it allows for direct comparison between datasets, which may be valuable for assessing reproducibility of results between palaeoproteomics experiments and laboratories.

As manuscripts typically focus on successful protocols and analyses, negative results are often not shared with the scientific community. This may lead to repeated unsuccessful attempts



at using a specific method, or analysing a particular type of sample, as researchers may not be aware of previous negative outcomes due to a lack of communication. In addition to wasting time and funding, this also risks a waste of archaeological samples, and may overinflate the image of how successful certain types of studies are. Often, publications featuring positive results can also publish associated negative results, such as a failure to assign taxonomy (Demeter *et al.* 2022), absence of protein preservation in a certain tissue (Chen *et al.* 2019) or at certain sites (Peters *et al.* 2023), or unexpectedly poor performance of a protease (Fagernäs *et al.* 2024b). Dissemination of negative results in the main text (as opposed to the supplementary information) will ensure that they reach the scientific community. Journals are increasingly willing to publish high-quality research that led to negative results, and specialised journals have also been founded purely for the purpose of reporting negative results (Brazil 2024). Negative results can also be shared at conferences, where they will reach a relevant audience, and associated conference abstracts may be published and become citable resources (e.g. Paleoanthropology Society & European Society for the study of Human Evolution 2024), although it should be noted that the amount of detail that can be shared is very limited, and as such, evaluating negative results is challenging. Additionally, preprints can be used to share reports on negative results, and empty or otherwise failed raw data files can be shared in data repositories. Finally, as colleagues, reviewers and editors, we can collectively avoid stigmatising non-significant test outcomes or negative results.

Alternatively, the recent development of pre-registering publications, known as Registered Reports, aims to circumvent the issues associated with publishing negative results (Chambers & Tzavella 2022). The details of what needs to be included in a Registered Report differ between journals, but the core concept is to submit a research plan including hypotheses, methodologies and planned analyses (Soderberg *et al.* 2021), which is subjected to peer review and can be either published separately or result in an in-principle acceptance of a later publication by a journal. This form of publication moves the peer-review to the start of the publication process (Lakens *et al.* 2024), so that an experiment is judged by the merits of its planning and negative results will not reduce opportunities for publication.

## Data Analysis and Interpretation

With improvements in protein extraction and mass spectrometry technology, and accompanying increased sizes of palaeoproteomic datasets, there is an increasing necessity of field-specific or adaptable software and pipelines for data analysis. In the wider proteomics community there is a call for increasing use of FAIR and open-source software (Perez-Riverol *et al.* 2024), and these principles should be applied also to palaeoproteomics analysis tools, to ensure transparency and replicability. Development, comparisons and benchmarking of software and pipelines can greatly benefit from a community-wide effort and communication, ensuring wide applicability and reliability across study materials, research groups and computing systems.

The analysis of palaeoproteomic data is fundamentally different between ZooMS peptide mass fingerprinting and shotgun proteomics experiments. ZooMS taxonomic identifications are often obtained by manually comparing the observed peaks to a reference list of biomarkers or

reference spectra. Some pre-processing steps of the spectra have become standard, such as averaging triplicate spectra, automatic peak picking and setting minimum signal-to-noise thresholds, but the association between an observed peak and a biomarker remains a manual step in most studies. Consequently, there is potential for a degree of subjectivity and inter-observer variability. To address this issue, there have been improvements in the development of automated tools for biomarker identification (Gu & Buckley 2018; Hickinbotham *et al.* 2020; Végh & Douka 2024). Such tools are of great interest and utility to the palaeoproteomics community, if the developed tools are made freely and openly available (e.g., <https://github.com/touzet/pampa/>). Reporting and interpretation guidelines have been established for ZooMS (Brown 2021) using mMass (Strohalm *et al.* 2010), which is currently the most popular tool for analysis of ZooMS data. It is a freely available and user-friendly software tool that easily allows the processing and display of MS1 spectra, but is no longer supported by its developer (though it remains accessible on GitHub: <https://github.com/xxao/mMass>). The development of a new open-access standard tool for visualising and processing ZooMS data would therefore be of great value to the palaeoproteomics community.

Of great importance to ZooMS analysis is the choice of biomarkers with which to compare the spectra. Since the initial publication of seven collagen peptide biomarkers for a small selection of taxa (Buckley *et al.* 2009), the repertoire of these biomarkers has been expanded to an increasingly wide range of species (Buckley *et al.* 2014; Speller *et al.* 2016; Welker *et al.* 2016). A key study in the biomarker development has been the standardisation of biomarker nomenclature and their explicit coupling to specific peptides rather than primarily being based on particular *m/z* windows (Brown *et al.* 2021a). This has facilitated the communication, discussion and validation of new biomarkers (Codlin *et al.* 2022; Dierickx *et al.* 2022; Janzen *et al.* 2021; Nel *et al.* 2023; Peters *et al.* 2021; Winter *et al.* 2023) and sets a standard against which proposed homologous biomarkers between taxa can be compared. Furthermore, the increasing number of species for which biomarkers are available has resulted in the rise of community-based and collaborative projects that aim to enhance accessibility, as well as ensure robustness through independent validation of biomarkers; something that the PAASTA community aims to play a key role in.

In contrast to the predominantly manual data analysis in ZooMS, shotgun proteomics experiments utilise protein search engines to match observed precursor and fragment ion spectra to peptide and protein databases. There are a large number of different protein search engines available, some of which are free to use - e.g., MaxQuant (Cox & Mann 2008), Fragpipe (Kong *et al.* 2017), and MetaMorpheus (Solntsev *et al.* 2018), while others require a paid license - e.g., Mascot (Perkins *et al.* 1999), PEAKS (Zhang *et al.* 2012), and Byonic (Bern *et al.* 2012). The number of options in protein search engines can make it overwhelming to make appropriate selections. Key features that help distinguish between different programmes are *de novo* sequencing and open searches. The usefulness of such features for data analysis depends greatly on the nature of the samples and the aim of the analysis. *De novo* sequencing is often applied in cases where it is known that there is no proteomic reference data available for all expected target species, and open searches are frequently applied when a high number of PTMs are expected. Beyond features such as open searches or *de novo* sequencing, there are more

subtle differences in the scoring algorithm used to identify peptides and proteins. Consequently, there are differences in the number of peptides and proteins that different protein search engines identify (Rodriguez Palomo *et al.* 2024), as well as which peptides and proteins they identify (Parker *et al.* 2021). It may therefore not always be possible to completely replicate the same peptide and protein identifications when analysing the same data using a different protein search engine. This also means that the use of proprietary software for protein analysis imposes a barrier to the reproducibility of the experiment. Furthermore, specific software may be more suited for the identification of short peptides, or high numbers of PTMs, and therefore comparing different search engines before deciding which one to use in the analysis may lead to improved protein and peptide identifications.

Pipelines for the analysis of shotgun proteomics datasets increase reproducibility and transparency in palaeoproteomics studies. Such pipelines have, for example, been developed for phylogenetic analyses based on palaeoproteomic data, including both the creation of a set of reference sequences translated from genomic sequences, as well as subsequent phylogenetic analysis (Patramanis *et al.* 2023). Further development of such pipelines, shared openly and including thorough documentation and tutorials, will help the entire palaeoproteomics community, as they can be applied to different datasets and be thoroughly tested by different researchers.

Apart from the choice of a particular protein search engine, the choice of a protein reference database has a large impact on the identification of peptides and proteins. The most straightforward effect is that any proteins not present in the database will not be identified in the sample. Instead, ions belonging to a protein present in the sample, but not in the protein reference database, may be erroneously assigned to a different protein or taxon (Knudsen & Chalkley 2011). Additionally, the relative size of the database, in terms of the number of protein sequences included, compared to the number of proteins present in the sample, can also impact protein identification (Rodriguez Palomo *et al.* 2024). This effect is described elsewhere (Li *et al.* 2016) in more detail, but perhaps the most significant implication of this phenomenon is that relatively small and constrained databases are preferred compared to broad databases. Despite this, the latter are still frequently used especially in the analysis of organic residues where little can be assumed regarding the composition of the sample. Finally, protein sequences available in reference databases can be incorrectly annotated (Harvey *et al.* 2021), which, if not caught, can cause issues for downstream interpretations. Considering the substantial effect of the composition of the database on the outcome of the analysis, many researchers create their own custom targeted databases. Although the use of custom databases has many advantages for protein analysis, it can complicate matters for those reviewing or interpreting the published data. If the database itself is not included in the publication it can be challenging to recreate it and, consequently, to replicate the analysis. Moreover, detailed metadata, including protein accession numbers, species names, and taxonomic identification numbers, are required to understand which proteins of which taxa are included in the database and therefore what taxa can or cannot be excluded as potential sources of the observed proteins. It is therefore recommended to either share the database or to provide highly detailed metadata for recreation of the database.

Considering the risk to reproducibility that comes with variety in protein analysis software, pipelines and databases, we reiterate earlier metadata reporting standards (Hendy *et al.* 2018b; Taylor & Goodlett 2005) in asking that any palaeoproteomic publication should strive for maximum openness regarding used parameters and at the minimum contain:

- Software, including version number, used to perform analysis.
- MS1 and MS2 search tolerances (i.e., precursor and fragment ion tolerance).
- Allowed fixed and variable PTMs and the maximum number of PTMs per peptide.
- Any cut off values, including peptide score, FDR (false discovery rate), and/or q-value.
- Minimum peptide length and when relevant maximum peptide length.
- Protease, or lack thereof, and protease specificity settings (e.g., specific, semispecific, missed cleavages, etc.)
- The type of search (i.e., open, error tolerant, *de novo*).
- Details regarding the database used for the analysis, including accession numbers and date of download, and optionally protein names, and taxonomic identifiers. Ideally the FASTA file should be made available, or otherwise all necessary metadata provided to replicate the database. Translated sequences with gene annotations, if available, should also be shared for species not covered in current protein databases.

To ensure reproducibility and transparency, it is also recommended to share information supporting protein and peptide identification, e.g. as part of a ProteomeXchange project or as supplementary files of a manuscript. For identified proteins, at the very least the accession number from the used database should be shared, to ensure that it is possible to trace the origin and additional information for the identification (such as protein name and taxonomy). Additionally, the number of supporting peptides and the peptides themselves should be shared. For identified peptides, the amino acid sequence and the assigned start and end positions within the protein are important information to share, as well as the number of supporting spectra, and any PTMs and their location. Sharing this information allows reuse of data without the need to re-generate protein and peptide identifications, as well as independent analysis of the results.

For ancient proteins, assessing their authenticity is essential, as they are often low-abundance and found in a mixture of proteins originating from a variety of sources (Hendy *et al.* 2018b). Reporting deamidation values for proteins or proteomes of interest, and comparing them with laboratory blanks or more recent samples, can aid in assessing the antiquity of the samples (Ramsøe *et al.* 2020). Additionally, processing and analysing laboratory blanks alongside samples will ensure that contamination from the laboratory or reagents are not interpreted as endogenous to the samples. As contamination has been shown to impact the reconstructed ancient proteome (Fagernäs *et al.* 2024a), decontamination with a protocol appropriate for each sample type is recommended prior to protein extraction. Finally, if, for example, dietary items are targeted in a study, analysing samples from the surrounding sediment or other tissues may ensure that environmental taxa are not mistakenly assumed to stem from the sample of interest (Mann *et al.* 2020).

Archaeological sciences are increasingly utilising code-based analysis methods tailored to specific research requirements, such as workflows written in R (R Core Team 2023) or Python.

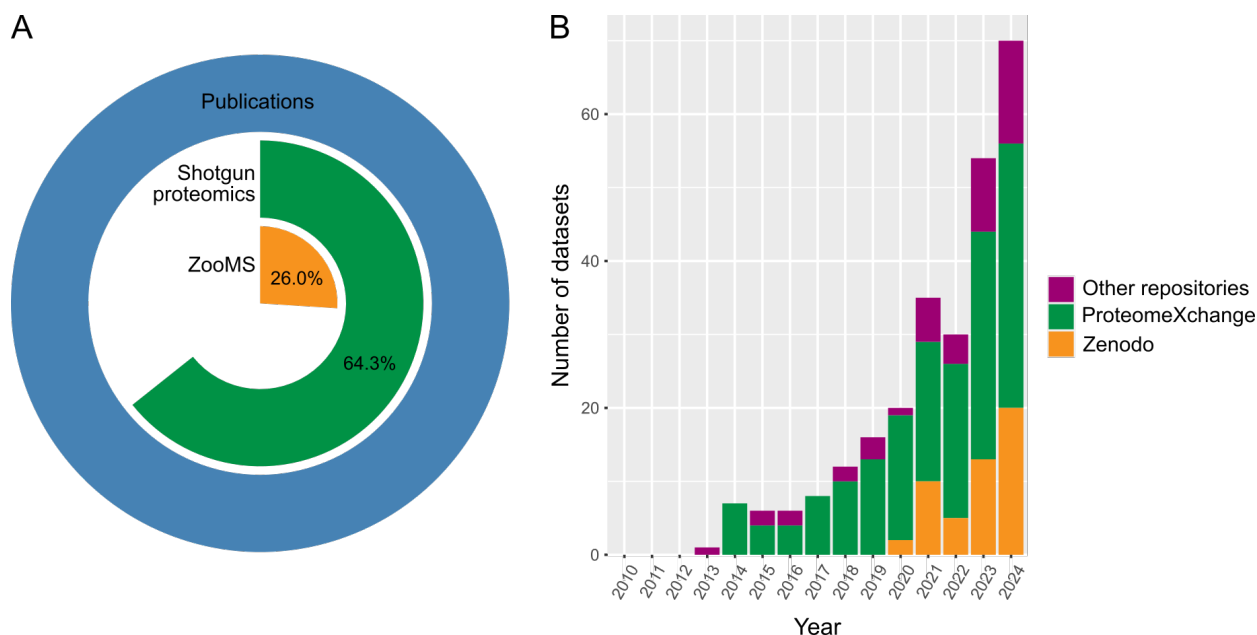
This allows for optimisation of analyses for each specific project's needs, automation of processes and standardisation of analysis workflows. Simultaneously, this requires code sharing for a study to be reproducible, and for it to be possible to thoroughly evaluate results. Methods sections in standard journals are often restricted in how much detail can be included. Code sharing with version tracking is possible through resources such as GitHub, and can be assigned a persistent identifier (e.g., digital object identifier, DOI). This can be done by, for example, archiving the code using Zenodo (e.g., Viñas-Caron *et al.* 2023, Zenodo DOI: 10.5281/zenodo.7406296). A study can thereby be fully transparent in precisely how essential steps, such as filtering of contaminant proteins or removal of low-quality peptides, were performed. This also allows for adaptation of analysis workflows written by other researchers, saving time and effort and allowing for citation of the original code document.

Although palaeoproteomic analyses can provide novel insights about our past, the proteomic results most often cannot stand on their own. For example, without dating and archaeological context, the taxonomic identification of a specimen cannot be meaningfully interpreted. It is therefore essential to collaborate with researchers from other fields, from the very beginning of a project, to ensure that all available information is combined into a holistic interpretation of the proteomic results. Archaeogenetic and palaeoproteomic analyses should not be seen as competing, but rather as complementary. For example, genetic sequences contain more phylogenetic information than protein sequences, but DNA degrades at a higher rate than some proteins, and the two fields therefore have different points of strength. They may also have differing strengths in the type of information they can provide. For example, in studies of archaeological dental calculus, genetic studies have been more promising for studying the oral microbiome (Fellows Yates *et al.* 2021; Ottoni *et al.* 2021; Warinner *et al.* 2014a), whereas proteomics has provided more information about past diets (Scott *et al.* 2021; Warinner *et al.* 2014a; Wilkin *et al.* 2021). Additionally, other emerging fields such as palaeometabolomics (Barberis *et al.* 2022; Brownstein *et al.* 2020; Huber *et al.* 2023; Velsko *et al.* 2017) can be integrated for a comprehensive biomolecular toolkit for analysis of archaeological materials.

## Data Sharing and Metadata Availability

A strength of ancient DNA research and a driver of growth of new software tools and interpretation was the consensus to make data openly available (Anagnostou *et al.* 2015). This development seems to still be in progress in the field of palaeoproteomics, whereby data is being produced and analysed but not always being made available alongside those publications (Fig. 2; Supplementary Dataset 1). This issue is most prominent for ZooMS studies, where raw data is only available for 26% of all studies that produced data (Fig. 2A). The picture is more positive for shotgun palaeoproteomics, with data being available for 64.3% of studies that produced data. Despite these low percentages, we do see an increase in publications with available data in recent years (Fig. 2B), most likely due to the concurrent increase in palaeoproteomics studies in recent years, and perhaps a response to an earlier call for the public sharing of raw and processed data by Hendy *et al.* (Hendy *et al.* 2018b). Although the problem of low reporting of raw data for ZooMS manuscripts has been previously discussed (Richter *et al.* 2022), there is currently no agreed-

upon standard for what data to publish, or how, alongside ZooMS manuscripts. We aim for the PAASTA community to facilitate the development of data publishing standards for both ZooMS and shotgun palaeoproteomics datasets, and we also hope to encourage researchers to revisit their prior work and, where possible, make that data available. It is worth noting that it is never too late to make data from prior publications available. We do, however, acknowledge that there might be factors that complicate open data sharing, for example, the importance of indigenous communities holding stewardship over their own data (Carroll *et al.* 2020).



**Figure 2. Data availability in palaeoproteomics.** A) Percentage of ZooMS and shotgun palaeoproteomics studies with available datasets. Preprints and papers that did not produce any data were excluded from this calculation. B) Number of palaeoproteomics datasets in online repositories (since 2010). ‘Other’ repositories include Dryad, Figshare, Github, Mendeley Data, and the OSF Framework. The dataset was compiled by manually going through the publications in Supplementary Dataset 1 and checking raw data availability. This figure was made in R v.4.4.2. with R Studio v.2024.12.0+467 (R Core Team 2023) using ggplot2 (Wickham 2016).

Equally important as open data sharing is the appropriate description of metadata. If the nature of the sample or the relationship between the analysed data and the samples is not clear, the reusability of data will decrease or disappear. The MAGE-TAB-Proteomics format (Dai *et al.* 2021) has been proposed by ProteomeXchange as a metadata format (Deutsch *et al.* 2023). MAGE-TAB-Proteomics has two components: the Investigation Description Format (IDF) and the Sample and Data Relationship Format (SDRF-Proteomics). The IDF, which is a general description of the research, is automatically generated from the mandatory information for registration in the ProteomeXchange repository, but the SDRF-Proteomics, which describes the details of the samples and the experiments, needs to be created and registered by the researcher. Since SDRF-Proteomics was not designed for archaeological materials, some information, such as the age or region of origin, specific to archaeological or paleontological materials, lacks a corresponding data input column. It may thus be necessary to create a revised format for

archeological/paleontological material-specific metadata, as has been necessary in the field of archaeogenetics (Bergström *et al.* 2024).

To aid in alleviating these concerns going forward and in an effort to establish best practices in the field, we propose the following as minimum reporting standards for palaeoproteomics, additionally adhering to the FAIR and CARE principles:

- When LC-MS/MS data is generated, the raw files should be made available in a public data repository. This should include raw data files from all analysed specimens, as well as extraction blanks. Our recommendation is to submit these files to ProteomeXchange, to which data can be deposited via several platforms including PRIDE and MassIVE. Submissions should ideally include the required files for a complete submission as detailed in their guidelines.
- Reconstructed protein sequences should be shared as FASTA files in supplementary data with the manuscript, or uploaded alongside raw data files to a repository such as ProteomeXchange. Note that GenBank and other protein data repositories do not accept submissions of reconstructed partial protein sequences derived from palaeoproteomics research, as there is no evidence for the length of the unknown sequence regions. Protein sequences should not be shared in .pdf format as this is a less accessible method of sharing.
- For MALDI-ToF or MALDI-FT-ICR data (for ZooMS), the raw data should be made available in a public data repository. Ideally, files with the processed data should also be made available for the sake of reproducibility. Our recommended repository for this data is Zenodo in .mzXML or .txt file format, or ProteomeXchange, as .mzXML files can be uploaded.
- Where code has been used to analyse data, the code should be provided so that analyses can be replicated and results/interpretations are transparent. We recommend sharing code through a version-tracked repository system such as GitHub or Zenodo, and we recommend that the code be archived using a persistent identifier, such as a DOI.
- Appropriate and relevant metadata of samples should be included in the form of a metadata file, such as SDRF-Proteomics, highlighting information such as:
  - Ancient and modern samples: links between samples and raw file names/spectra, sample IDs, batch and plate numbers, method of protein extraction, details of analytical machines used (including name of instrument and vendor) and precise parameters employed.
  - Ancient samples: context of the samples, dating method of samples and corresponding time period, notes of any treatments post-excavation.
  - Modern samples: source of reference materials, details of how species were identified, and if the sample was macerated, include notes on how this was done and how the sample has been stored and treated.
- All the shared files should be in an accessible format, from which data can be readily extracted (i.e., not PDFs, images, etc.).
- Ideally, research should be published in an Open Access format. However, we are also aware of the, at times, prohibitively expensive costs of publishing Open Access in numerous journals. There is an increasing number of Open Access journals, as well as

alternative publishing platforms, which can circumnavigate this issue. Additionally, it should be noted that authors are allowed to publicly archive the accepted version of a manuscript in institutional repositories. Ensuring that the corresponding data is made publicly available, even if the manuscript is not, is recommended and in spirit with our recommended best practices.

## Future Perspectives

The field of palaeoproteomics has advanced significantly over the past years, thanks to a range of emerging analytical and computational methodologies. Such innovations have enhanced the sensitivity and accuracy of ancient protein analysis, thereby providing deeper insights into the past. With these innovations leading to an ever-growing number of researchers and research applications, we are advocating to advance palaeoproteomics through keeping open science, collaboration and communication at the center of our research. These practices will enable our growing community of researchers to validate and build on each other's work, collaborate more effectively, and ultimately accelerate scientific progress in palaeoproteomics.

Future developments in both mass spectrometry and data processing are likely to further transform the field in a multitude of ways. Mass spectrometry instrumentation used in palaeoproteomics is crucial for advancing the field, although often accompanied by considerable costs and the need for proprietary software. Recently, new technologies for single-molecule proteomics that do not necessarily rely on mass spectrometry, such as nanopores or molecular markers, are being developed (Alfaro *et al.* 2021; Floyd & Marcotte 2022), and these new technologies have the potential to be applied to archaeological and paleontological samples (Alfaro *et al.* 2021). Frequent incremental upgrades in instrumentation and software or drastic technological innovation are often necessary to stay at the forefront of research, posing a potential barrier to widespread method adoption. However, continued investment in state-of-the-art instrumentation is essential for advancing palaeoproteomics research. AlphaFold, an artificial intelligence (AI) system developed by DeepMind (Jumper *et al.* 2021; Tunyasuvunakool *et al.* 2021), has revolutionised protein tertiary structure prediction, which may impact the future of palaeoproteomics. Additionally, a wide range of AI tools aimed at protein structure prediction have been introduced. In palaeoproteomic research, where recovered protein sequences are highly degraded, AlphaFold might aid researchers in inferring protein structure, thereby improving understanding in evolutionary relationships and functionality (Demarchi *et al.* 2022). Moreover, integrating AlphaFold predictions and reconstructed protein sequences might enhance the annotation and characterisation of ancient proteins. The use of AI in palaeoproteomics is likely to significantly advance the identification and analysis of proteins and proteomes, but will need to be extensively validated. These types of major advances can be efficiently evaluated and adapted into the field where appropriate, through collaboration between researchers from different subfields of palaeoproteomics.

Importantly, many of the challenges, opportunities and problems outlined above can be overcome by communication between researchers and research groups. This is true both within



the field of palaeoproteomics, as well as between research fields. Working together on projects (such as the creation of an open access, community-curated database of ZooMS peptide markers), software, and applications ensures that advances—and failures—are properly communicated. Such collaborative projects can also be used by the community to steer the direction of method development. Experiments that are valuable to the community, but are not easily funded by individual grants, such as reproducing previous results, validation of new methods or benchmarking tests, may become feasible if shared by many researchers from different institutions. Especially in the case of software validation, for which it can be prohibitively time consuming for the developer to test on a wide range of sample sets, a community-based approach may provide the large scale of diverse datasets needed and share the time-load required to validate new tools for palaeoproteomic applications.

Currently, a lot of palaeoproteomics expertise on the use of various extraction protocols, mass spectrometry settings and software is limited to a small number of individuals and institutions, and detailed (and practical) knowledge of palaeoproteomics is rarely taught at universities. As the palaeoproteomics community grows and the number of institutes with palaeoproteomics researchers increases, the importance of sharing knowledge and experiences increases as well. Not only will facilitating the sharing of technical know-how make palaeoproteomics more accessible to new researchers, it will also ensure that community conventions and high research standards are maintained, and provide researchers with an overview of the range of research tools available. Building a community such as PAASTA provides researchers with a platform for such conversations, allowing networking, collaboration, method development and validation, and knowledge sharing.

## Conclusions

The PAASTA community aims to create an open and supportive environment for palaeoproteomics researchers to share knowledge and expertise, and to foster collaborations and open science. We have written this manuscript as a means to establish and renew standards for best practice in palaeoproteomics research, emphasising data sharing and reporting, and promoting open science among the rapidly growing community of researchers within this field. Our key recommendations and proposed guidelines are as follows:

- Open and collaborative discussions with various stakeholders in cultural heritage regarding sample selection, documentation, planned proteomic analyses and interpretation.
- Detailed and reproducible protocols, as well as increased sharing of negative results..
- Development of analysis methods and pipelines in a community-supported manner, and detailed sharing of analysis methods and parameters.
- Raw data, protein databases, and relevant code to be made available alongside publications and in an accessible format and provide appropriate accompanying detailed metadata.

By adopting these guidelines of openness and collaboration, the palaeoproteomics community can enhance the quality, transparency and impact of its research. By jointly coming together from different research institutions, groups, and backgrounds we hope to be setting standards of best practices which will resonate with our community of researchers as a whole. PAASTA is an ever-evolving and growing community which welcomes all palaeoproteomics researchers to join. For more details we can be contacted at [paasta.community@gmail.com](mailto:paasta.community@gmail.com).

## Acknowledgements

We would like to thank the entire PAASTA community for discussions and collaborations which have led to this manuscript. This research has been made possible through funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement no. 787282 (Beasts2Craft, supporting M.C. and B.N.), no. 948365 (PROSPER, awarded to F.W.), and no. 865222 (EQuaTe, supporting M.R.D.), as well as under the Marie Skłodowska-Curie grant agreement no. 956351 (ChemArch, supporting J.D.), and the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101106627 (PROMISE, awarded to Z.F.), and ERC grant agreement no. 101041245 (MATRIX, supporting C.P.) and no. 101039416 (Fashioning Sudan, supporting L.C.V.C.), and the Volkswagen Foundation (C.W.), and by the Deutsche Forschungsgemeinschaft (460129525, NFDI4 Microbiota EnterArchaeo, C.W.), and the Werner Siemens Foundation (Paleobiotechnology, C.W.), and the U.S. National Science Foundation (C.W.), and the Harvard Radcliffe Institute (C.W.). T.T. is supported by the Japan Science and Technology Agency (FOREST, no. JPMJFR233D). R.M.W. is supported by the French National Research Agency (ANR) Young Researcher (JCJC) Grant awarded to Sophia V. Hansson (ANR-21-CE34-0001, ATCAF-project). F.W. is supported by VILLUM FONDEN (no. 40747). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, the European Research Council Executive Agency, or any of the other funders. Neither the European Union nor the granting agency can be held responsible for them.

## Code and data availability

R code and data have been archived at [10.5281/zenodo.14793421](https://doi.org/10.5281/zenodo.14793421).

## References

Abelson, P.H., 1954. Amino acids in fossils, *Science* 119(3096), 576.

Alfaro, J.A., P. Bohländer, M. Dai, M. Filius, C.J. Howard, X.F. van Kooten, S. Ohayon, A. Pomorski, S. Schmid, A. Aksimentiev, E.V. Anslyn, G. Bedran, C. Cao, M. Chinappi, E. Coyaoud, C. Dekker, G. Dittmar, N. Drachman, R. Eelkema, D. Goodlett, S. Hentz, U. Kalathiya, N.L. Kelleher, R.T. Kelly, Z. Kelman, S.H. Kim, B. Kuster, D. Rodriguez-Larrea, S. Lindsay, G. Maglia, E.M. Marcotte, J.P. Marino, C. Masselon, M. Mayer, P. Samaras, K. Sarthak, L. Sepiashvili, D. Stein, M. Wanunu, M. Wilhelm, P. Yin, A. Meller & C. Joo, 2021. The emerging landscape of single-molecule protein sequencing technologies, *Nature methods* 18(6), 604–17.

Anagnostou, P., M. Capocasa, N. Milia, E. Sanna, C. Battaggia, D. Luzi & G. Destro Bisol, 2015. When data sharing gets close to 100%: what human paleogenetics can teach the open science movement, *PloS one* 10(3), e0121409.

Ásmundsdóttir, R.D., J. Hansen, Z. Fagernäs, G. Troché, J.V. Olsen, M. Saña Seguí & F. Welker, 2024. Early Holocene preservation differences between cortical and trabecular bone proteomes, *Journal of archaeological science, reports* 57(104643), 104643.

Bache, N., P.E. Geyer, D.B. Bekker-Jensen, O. Hoerning, L. Falkenby, P.V. Treit, S. Doll, I. Paron, J.B. Müller, F. Meier, J.V. Olsen, O. Vorm & M. Mann, 2018. A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics, *Molecular & cellular proteomics: MCP* 17(11), 2284–96.

Barberis, E., M. Manfredi, E. Ferraris, R. Bianucci & E. Marengo, 2022. Non-invasive Paleo-metabolomics and Paleo-proteomics analyses reveal the complex funerary treatment of the early 18th Dynasty dignitary NEBIRI (QV30), *Molecules (Basel, Switzerland)* 27(21), 7208.

Bergström, A., J.A. Fellows Yates & C. Warinner, 2024. Ancient DNA data hold insights into past organisms and ecosystems - handle them with more care, *Nature* 636(8042), 296–98.

Bern, M., Y.J. Kil & C. Becker, 2012. Byonic: advanced peptide and protein identification software, *et al Current protocols in bioinformatics* 13(1), 13.20.1–13.20.14.

Brandt, L.Ø., M.R. Lillemark, M. Rytter, M.J. Collins & A.P. Tøttrup, 2022. Next Generation Lab puts Denmark's past in the hands of its future, *Antiquity* 96(385), 221–28.

Brazil, R., 2024. Illuminating 'the ugly side of science': fresh incentives for reporting negative results, *Nature*.

Brown, S., 2021. Identifying ZooMS Spectra (mammals) using mMass, *protocols.io*.

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

Brown, S., S. Hebestreit, N. Wang, N. Boivin, K. Douka & K. Korzow Richter, 2020. Zooarchaeology by Mass Spectrometry (ZooMS) for bone material - AmBiC protocol, *protocols.io*.

Brown, S., M. Kozlikin, M. Shunkov, A. Derevianko, T. Higham, K. Douka & K.K. Richter, 2021b. Examining collagen preservation through glutamine deamidation at Denisova Cave, *Journal of archaeological science* 133, 105454.

Brownstein, K.J., S. Tushingham, W.J. Damitio, T. Nguyen & D.R. Gang, 2020. An ancient residue metabolomics-based method to distinguish use of closely related plant species in ancient pipes, *Frontiers in molecular biosciences* 7, 133.

Buckley, M., 2015. Ancient collagen reveals evolutionary history of the endemic South American 'ungulates', *Proceedings. Biological sciences* 282(1806), 20142671.

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

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

Callaway, E., 2017. Stop hoarding ancient bones, plead archaeologists, *Nature*.

Cappellini, E., F. Welker, L. Pandolfi, J. Ramos-Madrugal, D. Samodova, P.L. R  ther, A.K. Fotakis, D. Lyon, J.V. Moreno-Mayar, M. Bukhsianidze, R. Rakownikow, Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, T.W. Stafford Jr, Y.L. Chan, A. G  therstr  m, S.K.S.S. Nathan, P.D. Heintzman, J.D. Kapp, I. Kirillova, Y. Moodley, J. Agusti, R.-D. Kahlke, G. Kiladze, B. Mart  nez-Navarro, S. Liu, M. Sandoval Velasco, M.-H.S. Sinding, C.D. Kelstrup, M.E. Allentoft, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dal  n, M.T.P. Gilbert, J.V. Olsen, D. Lordkipanidze & E. Willerslev, 2019. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny, *Nature* 574(7776), 103–7.

Carroll, S.R., I. Garba, O.L. Figueroa-Rodr  guez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, R. Sara, J.D. Walker, J. Anderson & M. Hudson, 2020. The CARE principles for indigenous data governance, *Data science journal* 19(1), .

Chambers, C.D. & L. Tzavella, 2022. The past, present and future of Registered Reports, *Nature human behaviour* 6(1), 29–42.

Chen, F., F. Welker, C.-C. Shen, S.E. Bailey, I. Bergmann, S. Davis, H. Xia, H. Wang, R. Fischer, S.E. Freidline, T.-L. Yu, M.M. Skinner, S. Stelzer, G. Dong, Q. Fu, G. Dong, J. Wang, D. Zhang & J.-J. Hublin, 2019. A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau, *Nature* 569(7756), 409–12.

Codlin, M.C., K. Douka & K.K. Richter, 2022. An application of zooms to identify archaeological avian fauna from Teotihuacan, Mexico, *Journal of archaeological science* 148, 105692.

Cox, J. & M. Mann, 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nature biotechnology* 26(12), 1367–72.

Dai, C., A. F  llgrabe, J. Pfeuffer, E.M. Solovyeva, J. Deng, P. Moreno, S. Kamatchinathan, D.J. Kundu, N. George, S. Fexova, B. Gr  ning, M.C. F  ll, J. Griss, M. Vaudel, E. Audain, M. Locard-Paulet, M. Turewicz, M. Eisenacher, J. Uszkoreit, T. Van Den Bossche, V. Schw  mmle, H. Webel, S. Schulze, D. Bouyss   , S. Jayaram, V.K. Duggineni, P. Samaras, M. Wilhelm, M. Choi, M. Wang, O. Kohlbacher, A. Brazma, I. Papatheodorou, N. Bandeira, E.W. Deutsch, J.A. Vizca  no, M. Bai, T. Sachsenberg, L.I. Levitsky & Y. Perez-Riverol, 2021. A proteomics sample metadata

representation for multiomics integration and big data analysis, *Nature communications* 12(1), 5854.

Demarchi, B., J. Stiller, A. Grealy, M. Mackie, Y. Deng, T. Gilbert, J. Clarke, L.J. Legendre, R. Boano, T. Sicheritz-Pontén, J. Magee, G. Zhang, M. Bunce, M.J. Collins & G. Miller, 2022. Ancient proteins resolve controversy over the identity of *Genyornis* eggshell, *Proceedings of the National Academy of Sciences of the United States of America* e2109326119.

Demeter, F., C. Zanolli, K.E. Westaway, R. Joannes-Boyau, P. Düringer, M.W. Morley, F. Welker, P.L. Rütger, M.M. Skinner, H. McColl, C. Gaunitz, L. Vinner, T.E. Dunn, J.V. Olsen, M. Sikora, J.-L. Ponche, E. Suzzoni, S. Frangeul, Q. Boesch, P.-O. Antoine, L. Pan, S. Xing, J.-X. Zhao, R.M. Bailey, S. Boualaphane, P. Sichanthongtip, D. Sihanam, E. Patole-Edoumba, F. Aubaile, F. Crozier, N. Bourgon, A. Zachwieja, T. Luangkoth, V. Souksavatdy, T. Sayavongkhamdy, E. Cappellini, A.-M. Bacon, J.-J. Hublin, E. Willerslev & L. Shackelford, 2022. A Middle Pleistocene Denisovan molar from the Annamite Chain of northern Laos, *Nature communications* 13(1), 2557.

Deutsch, E.W., N. Bandeira, Y. Perez-Riverol, V. Sharma, J.J. Carver, L. Mendoza, D.J. Kundu, S. Wang, C. Bandla, S. Kamatchinathan, S. Hewapathirana, B.S. Pullman, J. Wertz, Z. Sun, S. Kawano, S. Okuda, Y. Watanabe, B. MacLean, M.J. MacCoss, Y. Zhu, Y. Ishihama & J.A. Vizcaíno, 2023. The ProteomeXchange consortium at 10 years: 2023 update, *Nucleic acids research* 51(D1), D1539–48.

Dickinson, M.R., K. Scott, N.F. Adams, A.M. Lister & K.E.H. Penkman, 2024. Amino acid dating of Pleistocene mammalian enamel from the River Thames terrace sequence: a multi-taxon approach, *Quaternary geochronology* 82(101543), 101543.

Dierickx, K., S. Presslee, R. Hagan, T. Oueslati, J. Harland, J. Hendy, D. Orton, M. Alexander & V.L. Harvey, 2022. Peptide mass fingerprinting of preserved collagen in archaeological fish bones for the identification of flatfish in European waters, *Royal Society open science* 9(7), 220149.

Fagernäs, Z., M.I. García-Collado, J. Hendy, C.A. Hofman, C. Speller, I. Velsko & C. Warinner, 2020. A unified protocol for simultaneous extraction of DNA and proteins from archaeological dental calculus, *Journal of archaeological science* 118, 105135.

Fagernäs, Z., G. Troché, J.-P. Buylaert, T. Khujageldiev, R. Kurbanov, J.V. Olsen, M.W. Pedersen & F. Welker, 2024a. Cleaning the Dead: Optimized decontamination enhances palaeoproteomic analyses of a Pleistocene hominin tooth from Khudji, Tajikistan, *bioRxiv* 2024.06.13.598810.

Fagernäs, Z., G. Troché, J.V. Olsen & F. Welker, 2024b. Digging deeper into ancient skeletal proteomes through consecutive digestion with multiple proteases, *Journal of proteomics* 298, 105143.

Fellows Yates, J.A., I.M. Velsko, F. Aron, C. Posth, C.A. Hofman, R.M. Austin, C.E. Parker, A.E. Mann, K. Nägele, K.W. Arthur, J.W. Arthur, C.C. Bauer, I. Crevecoeur, C. Cupillard, M.C. Curtis, L. Dalén, M. Díaz-Zorita Bonilla, J.C. Díez Fernández-Lomana, D.G. Drucker, E. Escribano Escrivá, M. Francken, V.E. Gibbon, M.R. González Morales, A. Grande Mateu, K. Harvati, A.G. Henry, L. Humphrey, M. Menéndez, D. Mihailović, M. Peresani, S. Rodríguez Moroder, M. Roksandic, H. Rougier, S. Sázelová, J.T. Stock, L.G. Straus, J. Svoboda, B. Teßmann, M.J. Walker, R.C. Power, C.M. Lewis, K. Sankaranarayanan, K. Guschanski, R.W. Wrangham, F.E. Dewhurst, D.C. Salazar-García, J. Krause, A. Herbig & C. Warinner, 2021. The evolution and changing ecology of the African hominid oral microbiome, *Proceedings of the National Academy of Sciences of the United States of America* 118(20), .

Fleskes, R.E., A.C. Bader, K.S. Tsosie, J.K. Wagner, K.G. Claw & N.A. Garrison, 2022. Ethical guidance in human paleogenomics: New and ongoing perspectives, *Annual review of genomics and human genetics* 23(1), 627–52.

Floyd, B.M. & E.M. Marcotte, 2022. Protein sequencing, one molecule at a time, *Annual review of biophysics* 51(1), 181–200.

Fotakis, A.K., S.D. Denham, M. Mackie, M.I. Orbegozo, D. Mylopotamitaki, S. Gopalakrishnan, T. Sicheritz-Pontén, J.V. Olsen, E. Cappellini, G. Zhang, A. Christophersen, M.T.P. Gilbert & Å.J. Vågene, 2020. Multi-omic detection of *Mycobacterium leprae* in archaeological human dental calculus, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 375(1812), 20190584.

Giraldo, O., A. Garcia & O. Corcho, 2018. A guideline for reporting experimental protocols in life sciences, *PeerJ* 6, e4795.

Gu, M. & M. Buckley, 2018. Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting, *BMC bioinformatics* 19(1), 241.

Hansen, J., J. Dekker, G. Troché, Z. Fagernäs, J.V. Olsen, M. Saña Seguí & F. Welker, 2024. A comparative study of commercially available, minimally invasive, sampling methods on Early Neolithic humeri analysed via palaeoproteomics, *Journal of archaeological science* 167(106002), 106002.

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

Harvey, V.L., J.N. Keating & M. Buckley, 2021. Phylogenetic analyses of ray-finned fishes (Actinopterygii) using collagen type I protein sequences, *Royal Society Open Science* 8(8), 201955.

Haug, C., J.W.F. Reumer, J.T. Haug, A. Arillo, D. Audo, D. Azar, V. Baranov, R. Beutel, S. Charbonnier, R. Feldmann, C. Foth, R.H.B. Fraaije, P. Frenzel, R. Gašparič, D.E. Greenwalt, D. Harms, M. Hyžný, J.W.M. Jagt, E.A. Jagt-Yazykova, E. Jarzembowski, H. Kerp, A.G. Kirejtshuk, C. Klug, D.S. Kopylov, U. Kotthoff, J. Kriwet, L. Kunzmann, R.C. McKellar, A. Nel, C. Neumann, A. Nützel, V. Perrichot, A. Pint, O. Rauhut, J.W. Schneider, F.R. Schram, G. Schweigert, P. Selden, J. Szewo, B.W.M. van Bakel, T. van Eldijk, F.J. Vega, B. Wang, Y. Wang, L. Xing & M. Reich, 2020. Comment on the letter of the Society of Vertebrate Paleontology (SVP) dated April 21, 2020 regarding 'Fossils from conflict zones and reproducibility of fossil-based scientific data': the importance of private collections, *Palaontologische zeitschrift* 94(3), 413–29.

Hendy, J., 2021. Ancient protein analysis in archaeology, *Science advances* 7(3), .

Hendy, J., C. Warinner, A. Bouwman, M.J. Collins, S. Fiddyment, R. Fischer, R. Hagan, C.A. Hofman, M. Holst, E. Chaves, L. Klaus, G. Larson, M. Mackie, K. McGrath, A.Z. Mundorff, A. Radini, H. Rao, C. Trachsel, I.M. Velsko & C.F. Speller, 2018a. Proteomic evidence of dietary sources in ancient dental calculus, *Proceedings. Biological sciences / The Royal Society* 285(1883).

Hendy, J., F. Welker, B. Demarchi, C. Speller, C. Warinner & M.J. Collins, 2018b. A guide to ancient protein studies, *Nature ecology & evolution* 2(5), 791–99.

Hickinbotham, S., S. Fiddymment, T.L. Stinson & M.J. Collins, 2020. How to get your goat: automated identification of species from MALDI-ToF spectra, *Bioinformatics (Oxford, England)* 36(12), 3719–25.

Horsburgh, A.K., 2024. Concluding Commentary, in *Critical Perspectives on Ancient DNA*, eds. D. Strand, A. Källén & C. Mulcare. : The MIT Press, 167–75.

Huber, B., S. Hammann, C.E. Loebe, D.K. Jha, D.G. Vassão, T. Larsen, R.N. Spengler, D.Q. Fuller, P. Roberts, T. Deviese & N. Boivin, 2023. Biomolecular characterization of 3500-year-old ancient Egyptian mummification balms from the Valley of the Kings, *Scientific reports* 13(1), 12477.

Janzen, A., K.K. Richter, O. Mwebi, S. Brown, V. Onduso, F. Gatwiri, E. Ndiema, M. Katongo, S.T. Goldstein, K. Douka & N. Boivin, 2021. Distinguishing African bovids using Zooarchaeology by Mass Spectrometry (ZooMS): New peptide markers and insights into Iron Age economies in Zambia, *PloS one* 16(5), e0251061.

Jersie-Christensen, R.R., L.T. Lanigan, D. Lyon, M. Mackie, D. Belstrøm, C.D. Kelstrup, A.K. Fotakis, E. Willerslev, N. Lynnerup, L.J. Jensen, E. Cappellini & J.V. Olsen, 2018. Quantitative metaproteomics of medieval dental calculus reveals individual oral health status, *Nature communications* 9(1), 4744.

Jones, E.D., 2022. *Ancient DNA: The Making of a Celebrity Science*. : Yale University Press.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli & D. Hassabis, 2021. Highly accurate protein structure prediction with AlphaFold, *Nature*.

Källén, A., 2025. *The Trouble with Ancient DNA*. University of Chicago Press January 2025 . <https://press.uchicago.edu/ucp/books/book/chicago/T/bo238848353.html> [Accessed 20 Jan 2025].

Källén, A., C. Mulcare, A. Nyblom & D. Strand, 2024. Petrous fever: The gap between ideal and actual practice in ancient DNA research, *Current anthropology* 000–000.

Knudsen, G.M. & R.J. Chalkley, 2011. The effect of using an inappropriate protein database for proteomic data analysis, *PloS one* 6(6), e20873.

Kong, A.T., F.V. Lerepovost, D.M. Avtonomov, D. Mellacheruvu & A.I. Nesvizhskii, 2017. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics, *Nature methods* 14(5), 513–20.

Lakens, D., C. Mesquida, S. Rasti & M. Ditroilo, 2024. The benefits of preregistration and Registered Reports, *Evidence-Based Toxicology* 2(1), .

Le Meillour, L., V. Sinet-Mathiot, R.D. Ásmundsdóttir, J. Hansen, D. Mylopotamitaki, G. Troché, H. Xia, J. Herrera Bethencourt, K. Ruebens, G.M. Smith, Z. Fagernäs & F. Welker, 2024. Increasing sustainability in palaeoproteomics by optimizing digestion times for large-scale archaeological bone analyses, *iScience* 27(4), 109432.

Li, H., Y.S. Joh, H. Kim, E. Paek, S.-W. Lee & K.-B. Hwang, 2016. Evaluating the effect of database inflation in proteogenomic search on sensitive and reliable peptide identification, *BMC genomics* 17(Suppl 13), 1031.

Lowenstein, J.M., V.M. Sarich & B.J. Richardson, 1981. Albumin systematics of the extinct mammoth and Tasmanian wolf, *Nature* 291(5814), 409–11.

Mackie, M., P. R  ther, D. Samodova, F. Di Gianvincenzo, C. Granzotto, D. Lyon, D.A. Pegg  , H. Howard, L. Harrison, L.J. Jensen, J.V. Olsen & E. Cappellini, 2018. Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting, *Angewandte Chemie* 57(25), 7369–74.

Mann, A.E., J.A. Fellows Yates, Z. Fagern  s, R.M. Austin, E.A. Nelson & C.A. Hofman, 2020. Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus, *Quaternary international: the journal of the International Union for Quaternary Research*.

Miller, G.H., P.B. Beaumont, A.J. Jull & B. Johnson, 1992. Pleistocene geochronology and palaeothermometry from protein diagenesis in ostrich eggshells: implications for the evolution of modern humans, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 337(1280), 149–57.

Nair, B., I.R. Palomo, B. Markussen, C. Wiuf, S. Fiddym  nt & M.J. Collins, 2023. Parchment Glutamine Index (PQI): A novel method to estimate glutamine deamidation levels in parchment collagen obtained from low-quality MALDI-TOF data, *Peer community journal* 3(e10), e10.

Nel, T.H., C. Peters, K.K. Richter, C. Henshilwood, K. van Niekerk & K. Douka, 2023. Peptide mass fingerprinting as a tool to assess micromammal biodiversity in Pleistocene South Africa: The case of Klipdrift Shelter, *Quaternary science reviews* 322, 108380.

Otoni, C., D. Bori  , O. Cheronet, V. Sparacello, I. Dori, A. Coppa, D. Antonovi  , D. Vujevi  , T. Douglas Price, R. Pinhasi & E. Cristiani, 2021. Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus, *Proceedings of the National Academy of Sciences of the United States of America* 118(32), .

Paleoanthropology Society & European Society for the study of Human Evolution, 2024. 14th Annual Meeting of the European Society for the study of Human Evolution Abstracts Zagreb, 11-15 September 2024, *PalaeoAnthropology* 2, 365–558.

P  lsd  ttir, A.H., A. Bl  uer, E. Rannam  e, S. Boessenkool & J.H. Hallsson, 2019. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains, *Royal Society open science* 6(10), 191059.

Parker, R., A. Tailor, X. Peng, A. Nicastr  , J. Zerweck, U. Reimer, H. Wenschuh, K. Schnatbaum & N. Ternette, 2021. The choice of search engine affects sequencing depth and HLA class I allele-specific peptide repertoires, *Molecular & cellular proteomics: MCP* 20(100124), 100124.

Paterson, R.S., P.P. Madupe & E. Cappellini, 2024. Paleoproteomics sheds light on million-year-old fossils, *Nature reviews. Molecular cell biology* 26(1), 1–2.

Patramanis, I., J. Ramos-Madrigal, E. Cappellini & F. Racimo, 2023. PaleoProPhyler: a reproducible pipeline for phylogenetic inference using ancient proteins, *Peer Community Journal* 3, .



Penkman, K.E.H., R.C. Preece, D.R. Bridgland, D.H. Keen, T. Meijer, S.A. Parfitt, T.S. White & M.J. Collins, 2011. A chronological framework for the British Quaternary based on Bithynia opercula, *Nature* 476(7361), 446–49.

Perez-Riverol, Y., W. Bittremieux, W.S. Noble, L. Martens, A. Bilbao, M.R. Lazear, B. Grüning, D.S. Katz, M.J. MacCoss, C. Dai, J.K. Eng, R. Bouwmeester, M.R. Shortreed, E. Audain, T. Sachsenberg, J. Van Goey, G. Wallmann, B. Wen, L. Käll & W.E. Fondrie, 2024. Open-source and FAIR Research Software for Proteomics, *ChemRxiv*.

Perkins, D.N., D.J. Pappin, D.M. Creasy & J.S. Cottrell, 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis* 20(18), 3551–67.

Peters, C., K.K. Richter, T. Manne, J. Dortch, A. Paterson, K. Travouillon, J. Louys, G.J. Price, M. Petraglia, A. Crowther & N. Boivin, 2021. Species identification of Australian marsupials using collagen fingerprinting, *Royal Society open science* 8(10), 211229.

Peters, C., Y. Wang, V. Vakil, J. Cramb, J. Dortch, S. Hocknull, R. Lawrence, T. Manne, C. Monks, G.E. Rössner, H. Ryan, M. Siversson, T. Ziegler, J. Louys, G.J. Price, N. Boivin & M.J. Collins, 2023. Bone collagen from subtropical Australia is preserved for more than 50,000 years, *Communications Earth & Environment* 4(1), 1–8.

Poujois, P. & L. Le Meillour, 2023. Palaeoproteomics protocol - arid environment samples, *protocols.io*.

Presslee, S., K. Penkman, R. Fischer, E. Richards-Slidel, J. Southon, C.A. Hospitaleche, M. Collins & R. MacPhee, 2021. Assessment of different screening methods for selecting palaeontological bone samples for peptide sequencing, *Journal of proteomics* 230(103986), 103986.

Ramsøe, A., V. van Heekeren, P. Ponce, R. Fischer, I. Barnes, C. Speller & M.J. Collins, 2020. DeamIDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes, *Journal of archaeological science* 115, 105080.

Rayfield, E.J., J.M. Theodor & P.D. Polly, 2020. *Fossils from Conflict Zones and Reproducibility of Fossil-Based Scientific Data*.

R Core Team, 2023. *R: A Language and Environment for Statistical Computing*.

Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019. *CARE Principles*. GIDA-global.org 2019 . <https://www.gida-global.org/care> [Accessed 28 Nov 2024].

Richter, K.K., M. Codlin, M. Seabrook & C. Warinner, 2022. A primer for ZooMS applications in archaeology, *Proceedings of the National Academy of Sciences* 119(20), e2109323119.

Rodriguez Palomo, I., B. Nair, Y. Chiang, J. Dekker, B. Dartigues, M. Mackie, M. Evans, R. Macleod, J.V. Olsen & M.J. Collins, 2024. Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins, *Peer community journal* 4(e107), .

Rusu, I., I. Paica, A. Vulpoi, C. Radu, C. Mircea, C. Dobrinescu, V. Bodolică & B. Kelemen, 2019. Dual DNA-protein extraction from human archeological remains, *Archaeological and*

*anthropological sciences* 11(7), 3299–3307.

Rüther, P.L., I.M. Husic, P. Bangsgaard, K.M. Gregersen, P. Pantmann, M. Carvalho, R.M. Godinho, L. Friedl, J. Cascalheira, A.J. Taurozzi, M.L.S. Jørvik, M.M. Benedetti, J. Haws, N. Bicho, F. Welker, E. Cappellini & J.V. Olsen, 2022. SPIN enables high throughput species identification of archaeological bone by proteomics, *Nature communications* 13(1), 2458.

Saitta, E.T., J. Vinther, M.K. Crisp, G.D. Abbott, L. Wheeler, S. Presslee, T.G. Kaye, I. Bull, I. Fletcher, X. Chen, D. Vidal, F. Sanguino, Á.D. Buscalioni, J. Calvo, P.C. Sereno, S.L. Baumgart, M. Pittman, M.J. Collins, J. Sakalauskaite, M. Mackie, F.D. Bello, M.R. Dickinson, M.A. Stevenson, P. Donohoe, P.R. Heck, B. Demarchi & K.E.H. Penkman, 2024. Non-avian dinosaur eggshell calcite can contain ancient, endogenous amino acids, *Geochimica et cosmochimica acta* 365, 1–20.

Santucci, V.L., P.B. Newman & B.D. Taff, 2016. Toward a conceptual framework for assessing the human dimensions of paleontological resources, *Fossil Record 5, New Mexico Museum of Natural History and Science Bulletin* 74 239–48.

Scott, A., R.C. Power, V. Altmann-Wendling, M. Artzy, M.A.S. Martin, S. Eisenmann, R. Hagan, D.C. Salazar-García, Y. Salmon, D. Yegorov, I. Milevski, I. Finkelstein, P.W. Stockhammer & C. Warinner, 2021. Exotic foods reveal contact between South Asia and the Near East during the second millennium BCE, *Proceedings of the National Academy of Sciences of the United States of America* 118(2), .

Scott, A., S. Reinhold, T. Hermes, A.A. Kalmykov, A. Belinskiy, A. Buzhilova, N. Berezina, A.R. Kantorovich, V.E. Maslov, F. Guliyev, B. Lyonnet, P. Gasimov, B. Jalilov, J. Eminli, E. Iskandarov, E. Hammer, S.E. Nugent, R. Hagan, K. Majander, P. Onkamo, K. Nordqvist, N. Shishlina, E. Kaverzneva, A.I. Korolev, A.A. Khokhlov, R.V. Smolyaninov, S.V. Sharapova, R. Krause, M. Karapetian, E. Stolarczyk, J. Krause, S. Hansen, W. Haak & C. Warinner, 2022. Emergence and intensification of dairying in the Caucasus and Eurasian steppes, *Nature ecology & evolution* 6(6), 813–22.

Scott, A. & C. Warinner, 2020. Ancient Proteins Extraction Protocol, *protocols.io*.

Soderberg, C.K., T.M. Errington, S.R. Schiavone, J. Bottesini, F.S. Thorn, S. Vazire, K.M. Esterling & B.A. Nosek, 2021. Initial evidence of research quality of registered reports compared with the standard publishing model, *Nature human behaviour* 5(8), 990–97.

Solntsev, S.K., M.R. Shortreed, B.L. Frey & L.M. Smith, 2018. Enhanced Global Post-translational Modification Discovery with MetaMorpheus, *Journal of proteome research* 17(5), 1844–51.

Somel, M., N.E. Altınışık, F. Özer & M.C. Ávila-Arcos, 2021. Collaborate equitably in ancient DNA research and beyond, *Nature* 600(7887), 37.

Speller, C., Y. van den Hurk, A. Charpentier, A. Rodrigues, A. Gardeisen, B. Wilkens, K. McGrath, K. Rowsell, L. Spindler, M. Collins & M. Hofreiter, 2016. Barcoding the largest animals on Earth: ongoing challenges and molecular solutions in the taxonomic identification of ancient cetaceans, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371(1702), .

Standvoss, K., V. Kazezian, B.R. Lewke, K. Bastian, S. Chidambaram, S. Arafat, U. Alsharif, A. Herrera-Melendez, A.-D. Knipper, B.M.S. Seco, N.N. Soto, O. Rakitzis, I. Steinecker, P. van

Kronenberg Till, F. Zarebidaki, P. Abbasi & T.L. Weissgerber, 2024. Shortcut citations in the methods section: Frequency, problems, and strategies for responsible reuse, *PLoS biology* 22(4), e3002562.

Stewart, N.A., R.F. Gerlach, R.L. Gowland, K.J. Gron & J. Montgomery, 2017. Sex determination of human remains from peptides in tooth enamel, *Proceedings of the National Academy of Sciences of the United States of America* 114(52), 13649–54.

Strohalm, M., D. Kavan, P. Novák, M. Volný & V. Havlíček, 2010. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data, *Analytical chemistry* 82(11), 4648–51.

Taylor, G.K. & D.R. Goodlett, 2005. Rules governing protein identification by mass spectrometry, *Rapid communications in mass spectrometry: RCM* 19(23), 3420.

Tunyasuvunakool, K., J. Adler, Z. Wu, T. Green, M. Zielinski, A. Židek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G.J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S.A.A. Kohl, A. Potapenko, A.J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A.W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper & D. Hassabis, 2021. Highly accurate protein structure prediction for the human proteome, *Nature* 596(7873), 590–96.

Végh, E.I. & K. Douka, 2024. SpecieScan: Semi-Automated taxonomic identification of bone collagen peptides from MALDI-ToF MS spectra, *Bioinformatics* 40(3), btae054.

Velsko, I.M., K.A. Overmyer, C. Speller, L. Klaus, M.J. Collins, L. Loe, L.A.F. Frantz, K. Sankaranarayanan, C.M. Lewis Jr, J.B.R. Martinez, E. Chaves, J.J. Coon, G. Larson & C. Warinner, 2017. The dental calculus metabolome in modern and historic samples, *Metabolomics: Official journal of the Metabolomic Society* 13(11), 134.

Venable, J.D., M.-Q. Dong, J. Wohlschlegel, A. Dillin & J.R. Yates, 2004. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra, *Nature methods* 1(1), 39–45.

Viñas-Caron, L.C., I. Rodríguez Palomo, N. Fazlic, J. Vnouček, M. Driscoll, S. Fiddymment & M.J. Collins, 2023. A biological reading of a palimpsest, *iScience* 26(6), 106786.

Warinner, C., J. Hendy, C. Speller, E. Cappellini, R. Fischer, C. Trachsel, J. Arneborg, N. Lynnerup, O.E. Craig, D.M. Swallow, A. Fotakis, R.J. Christensen, J.V. Olsen, A. Liebert, N. Montalva, S. Fiddymment, S. Charlton, M. Mackie, A. Canci, A. Bouwman, F. Rühli, M.T.P. Gilbert & M.J. Collins, 2014a. Direct evidence of milk consumption from ancient human dental calculus, *Scientific reports* 4, 7104.

Warinner, C., K. Korzow Richter & M.J. Collins, 2022. Paleoproteomics, *Chemical reviews* 122(16), 13401–46.

Warinner, C., J.F.M. Rodrigues, R. Vyas, C. Trachsel, N. Shved, J. Grossmann, A. Radini, Y. Hancock, R.Y. Tito, S. Fiddymment, C. Speller, J. Hendy, S. Charlton, H.U. Luder, D.C. Salazar-García, E. Eppler, R. Seiler, L.H. Hansen, J.A.S. Castruita, S. Barkow-Oesterreicher, K.Y. Teoh, C.D. Kelstrup, J.V. Olsen, P. Nanni, T. Kawai, E. Willerslev, C. von Mering, C.M. Lewis Jr, M.J. Collins, M.T.P. Gilbert, F. Rühli & E. Cappellini, 2014b. Pathogens and host immunity in the ancient human oral cavity, *Nature genetics* 46(4), 336–44.

Wehmiller, J.F., P.E. Hare & G.A. Kujala, 1976. Amino acids in fossil corals: racemization (epimerization) reactions and their implications for diagenetic models and geochronological studies, *Geochimica et cosmochimica acta* 40(7), 763–76.

Welker, F., M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J.V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.-J. Hublin, L. Orlando, I. Barnes & R.D.E. MacPhee, 2015. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, *Nature* 522(7554), 81–84.

Welker, F., M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J.R. Stewart, S. Pääbo, M.J. Collins & J.-J. Hublin, 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, *Proceedings of the National Academy of Sciences of the United States of America* 113(40), 11162–67.

Welker, F., J. Ramos-Madrigal, P. Gutenbrunner, M. Mackie, S. Tiwary, R. Rakownikow, Jersie-Christensen, C. Chiva, M.R. Dickinson, M. Kuhlwilm, M. de Manuel, P. Gelabert, M. Martínón-Torres, A. Margvelashvili, J.L. Arsuaga, E. Carbonell, T. Marques-Bonet, K. Penkman, E. Sabidó, J. Cox, J.V. Olsen, D. Lordkipanidze, F. Racimo, C. Lalueza-Fox, J.M. Bermúdez de Castro, E. Willerslev & E. Cappellini, 2020. The dental proteome of *Homo antecessor*, *Nature* 580(7802), 235–38.

Wickham, H., 2016. *ggplot2*. : Springer New York.

Wilkin, S., L.T. Lanigan, N. Montes, M. Sharma, C. Avanzi, D. Sejdiu, K. Majander, S. Pfrengle, Y. Chiang, L. Kunz, A. Dittmann, F. Rühli, P. Singh, M.F. Coll, M.J. Collins, A.J. Taurozzi & V.J. Schuenemann, 2024. Sequential trypsin and ProAlanase digestions unearth immunological protein biomarkers shrouded by skeletal collagen, *iScience* 27(5), 109663.

Wilkinson, M.D., M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao & B. Mons, 2016. The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data* 3(1), 160018.

Wilkin, S., A. Ventresca Miller, R. Fernandes, R. Spengler, W.T.-T. Taylor, D.R. Brown, D. Reich, D.J. Kennett, B.J. Culleton, L. Kunz, C. Fortes, A. Kitova, P. Kuznetsov, A. Epimakhov, V.F. Zaiibert, A.K. Outram, E. Kitov, A. Khokhlov, D. Anthony & N. Boivin, 2021. Dairying enabled Early Bronze Age Yamnaya steppe expansions, *Nature* 598(7882), 629–33.

Wilson, J., N.L. van Doorn & M.J. Collins, 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen, *Analytical chemistry* 84(21), 9041–48.

Winter, R.M., W. de Kock, M. Mackie, M. Ramsøe, E. Desiderà, M. Collins, P. Guidetti, S. Presslee, M.M. Alegre, T. Oueslati, A.M. Muniz, D. Michailidis, Y. van den Hurk, A.J. Taurozzi & C. Çakirlar, 2023. Grouping groupers in the Mediterranean: Ecological baselines revealed by

ancient proteins, *Ecology and evolution* 13(10), e10625.

Zhang, J., L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G.A. Lajoie & B. Ma, 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification, *Molecular & cellular proteomics: MCP* 11(4), M111.010587.