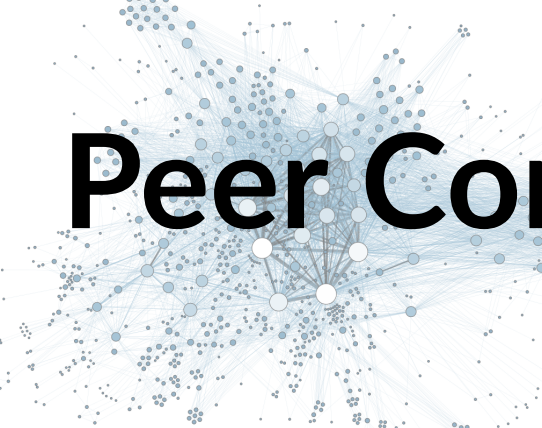




Peer Community Journal

Section: Archaeology

Opinion / Perspective

Published
2025-08-06

Cite as

Joannes Adrianus Antonius Dekker, Carli Peters, Rachel M. Winter, Matthew J. Collins, Marc R. Dickinson, Virginia L. Harvey, Eden Hill, Bharath Nair, Takumi Tsutaya, Laura C. Viñas-Caron, Christina Warinner, Frido Welker and Zandra Fagernäs (2025)
Open science, communication, and collaboration for the future of palaeoproteomics, Peer Community Journal, 5: XX.

Correspondence

zandra.fagernas@sund.ku.dk

Peer-review

Peer reviewed and
recommended by
PCI Archaeology,

<https://doi.org/10.24072/pci.archaeo.100601>



This article is licensed
under the Creative Commons
Attribution 4.0 License.

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

Joannes Adrianus Antonius Dekker^{1,2,3}, Carli Peters^{4,5}, Rachel M. Winter⁶, Matthew J. Collins⁷, Marc R. Dickinson⁸, Virginia L. Harvey⁹, Eden Hill^{10,11}, Bharath Nair¹², Takumi Tsutaya¹³, Laura C. Viñas-Caron¹⁴, Christina Warinner¹⁵, Frido Welker¹⁶, and Zandra Fagernäs¹⁷

Volume 5 (2025), article XX

<https://doi.org/10.24072/pcjournal.576>

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.

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

Peer Community Journal is a member of the
Centre Mersenne for Open Scientific Publishing
<http://www.centre-mersenne.org/>

e-ISSN 2804-3871



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 (Figure 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.

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 (Warinner et al., 2014a; Hendy et al., 2018a; Scott et al., 2022). Similarly, proteins deriving from pathogens or the immune response can give indications of the presence of a disease (Warinner et al., 2014b; Fotakis et al., 2020; 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 remains even when no morphologically identifiable characteristics are present, or when osteomorphology is too similar between taxa for distinctions. This approach has also been successful in the study of cloth materials like textiles, fur and leather (Hollemeyer et al., 2008; Brandt et al., 2014; Az  mard et al., 2021; Solazzo & Phipps, 2022; Vi  nas-Caron & Brandt, 2025), as well as art and heritage objects such as paintings and manuscripts (Kirby et al., 2013, 2020, 2023; Fiddyment et al., 2015, 2019; Teasdale et al., 2017; Di Gianvincenzo et al., 2023). Furthermore, variation in amino acid sequences between taxa also allows for phylogenetic placement (Lowenstein et al., 1981; Buckley, 2015; Welker et al., 2015, 2020; Cleland et al., 2015, 2016; Schroeter et al., 2017; Cappellini et al., 2019; Harvey et al., 2021; Demarchi et al., 2022), 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 proteins. For example, deamidation of asparagine and glutamine has been proposed as both a preservation or degradation indicator and a means of assessing the relative preservation of a detected protein or peptide (Wilson et al., 2012; Cleland et al., 2015; Schroeter & Cleland, 2016; Rams  e et al., 2020, 2021; Brown et al., 2021a), 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). However, caution should be applied when interpreting deamidation for authentication purposes, as factors other than age can also influence deamidation rates (Brown et al., 2021a). 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 (Wehmiller et al., 1976; Miller et al., 1992; Penkman et al., 2011; Dickinson et al., 2024). 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).

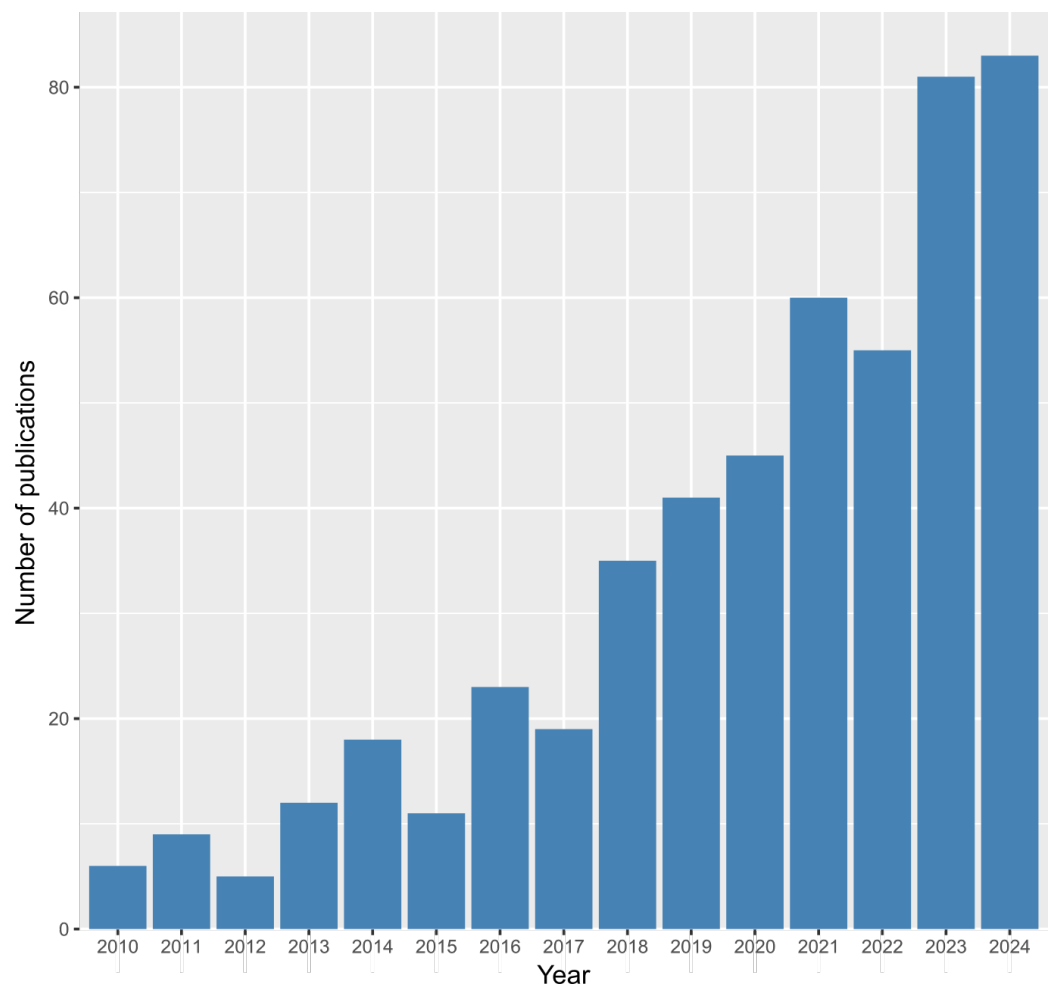


Figure 1 - Number of palaeoproteomics articles published each year since 2010 (excluding preprints). The list of papers was compiled by a Web of Science query (accessed on 14 June 2024) using the search terms: “palaeoproteomics” OR “paleoproteomics” OR “ancient proteins” OR “ZooMS” OR “collagen fingerprinting” 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 2009). See also the archived Dataset on Zenodo (<https://doi.org/10.5281/zenodo.14793421>) for a full record of the curated dataset.

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 (Hendy et al. 2018b; Fagernäs et al., 2025). 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 in other fields, especially for early career researchers, through competition and lack of communication (Callaway, 2017; Jones, 2022; Horsburgh, 2024; 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. Depending on the context of the material being studied, ethical sampling guidelines by discipline may already be in place and available for consultation (e.g. Institute of Conservation Heritage Science Group (Quye & Strlič, 2019)), Paleontological Research Institution's Special Terms and Conditions for Technical Analysis/Destructive Sampling). Such documents can provide palaeoproteomics researchers with initial insight into the perspectives and guidelines of their collaborators. 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;

(Haghighi et al., 2024)) and manufacturing processes (e.g. liming, tanning, dyeing; (Solazzo et al., 2014; Ebsen et al., 2019; Nair et al., 2023)). 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, thoroughly discussed and understood by all parties in advance, and that additional research opportunities have been carefully considered. Wherever possible, the use of ‘self-sampled’ material—fragments that have already detached from the object or material—should be prioritised to avoid damage to the main mass. Conversations supported by visual aids (e.g., Kirby et al., 2020) demonstrating the minimal sample sizes and minimally destructive approaches used in palaeoproteomics (Fiddymment et al., 2015) can help illustrate how the field is striving to balance information gain with the preservation of cultural and natural heritage materials. There is no single right answer for this tradeoff between preservation and obtaining new information through (potentially) destructive analysis, rather it will be extremely case-specific, depending on the amount of material available, the research question and the risk of an unsuccessful analysis. The core of conversations with curators and other stakeholders should therefore be to obtain ‘informed consent’, with all parties subscribing to the aim of the analysis and being aware of the risks involved. Given that the methods for extracting aDNA, proteins, and lipids target different physicochemical characteristics of each material, it should be the intention to sample in a way that maximises the potential to generate complementary datasets from the same sub-samples (e.g., for archaeogenetic analysis, isotopic analysis, radiocarbon dating).

Palaeoproteomic analyses are destructive to varying degrees, as part of a specimen is irreversibly removed, but the degree of destructiveness varies depending on the chosen extraction protocol. Minimally destructive protocols have also been developed, but vary in their success levels (Evans et al., 2023; 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 (Fiddymment et al., 2015; Pálsdóttir et al., 2019; Fleskes et al., 2022). This will also aid in understanding the conservation history of the sample, which may influence which analyses can be conducted, such as the introduction of modern contaminants or treatment with animal-based glues (van der Sluis et al., 2023; Haghighi et al., 2024). 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 should 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. Palaeoproteomic research offers the opportunity to open up

new lines of questioning and address research questions that other methodologies are not well suited for, but its full value can only be achieved through integration in theoretical frameworks including data obtained from other methodologies. 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 (Sinet-Mathiot et al., 2019, 2023; 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 (Rusu et al., 2019; Fagernäs et al., 2020). Where possible, any remaining material following analyses should be carefully returned to collections to support future research opportunities and reduce the need for additional sampling. 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. While a single sample may be enough for taxonomic identification of a specimen, other applications such as method development and comparisons between populations or archaeological sites need more samples, and it is essential to ensure that a statistically robust experimental design is set up prior to sampling. We also want to emphasise the necessity of equality between fields in a multidisciplinary study. All methods, including biomolecular techniques such as palaeoproteomics, but also macro- or microscopic methodologies, have their own limitations and advantages. The added value of multidisciplinary lies in the integration of the contributions from different analyses in a single framework.

The rights of Indigenous peoples and the colonial nature of resource-intensive ancient biomolecular research have long been debated in archaeology and archaeological science, and most recently heavily discussed in the field of archaeogenetics (Somel et al., 2021; Fleskes et al., 2022; Källén et al., 2024; Källén, 2025). As an emerging field, palaeoproteomics can learn from the discussion on the ethics of archaeological 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 aDNA studies (Harvey et al., 2018; Cappellini et al., 2019; Welker et al., 2020; Demarchi et al., 2022), the ethical dimensions of palaeontological research, which have not been much discussed in aDNA research, may become a challenge for palaeoproteomics. For example, palaeontological specimens are sometimes commercially traded or privately owned. Although more rarely the case in archaeology, similar ethical concerns arise in the study of privately owned or commercially traded art objects, such as paintings, where issues of provenance, access, and research use must also be carefully considered. The research use of such specimens is debated in palaeontology (Santucci et al., 2016; Rayfield et al., 2020; Haug et al., 2020), 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, legal and ownership complexities surrounding these materials. Existing frameworks, such as the Society for Vertebrate Paleontology's ethical guidelines regarding the use of commercial specimens (Society of Vertebrate Paleontology, 2025), may provide useful models for developing such standards.

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; Scott & Warinner, 2020; Poujois & Le Meillour, 2023). 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. This would ensure both that new developments are also considering the needs of palaeoproteomics, as well as that palaeoproteomics researchers are aware of the limitations of the applicability of new methodology to their research. 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 the development of bead-based protein capture and digestion approaches, EvoSep-based liquid chromatography workflows (Bache et al., 2018), and DIA data acquisition strategies (Venable et al., 2004). These developments are primarily aimed for clinical proteomics, but are also suitable for the analysis of ancient proteins. In the context of palaeoproteomics, efforts to standardise LC-MS/MS workflows, such as the adoption of systems with predefined, reproducible chromatographic settings, can facilitate direct comparison between datasets and enhance reproducibility across experiments and laboratories. Although the use of fixed chromatographic methods limits the degree of optimisation for specific sample types, it promotes

consistency, which is critical for building comparable palaeoproteomic datasets across the research community.

As manuscripts typically focus on successful protocols and analyses, negative results are often not shared with the scientific community. We recognise that this may be due to pressure to publish polished high-impact studies, however, avoiding dissemination of negative results is in the end detrimental to the field as a whole. 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., 2024). Mentioning negative results in the main text (as opposed to only being included in 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., 2025), 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; Hickenbotham et al., 2020; Baker et al., 2023; 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; Xia et al., 2024). An overview of all available biomarkers can be found in a Google Spreadsheet kept by the University of York (University of York, 2025). This is a live document that is continuously being updated as new biomarkers are published. 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., 2021b). This has facilitated the communication, discussion and validation of new biomarkers (Janzen et al., 2021; Peters et al., 2021; Dierickx et al., 2022; Codlin et al., 2022; Winter et al., 2023; Nel 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), MetaMorpheus (Solntsev et al., 2018) and pFind (Chi 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). Additionally, a pipeline has been written for collagen-based taxonomic classification through LC-MS/MS data, which is also capable of identifying mixtures of species in a sample (Engels et al., 2025). 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 (Sticker et al., 2017). 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 (Taylor & Goodlett, 2005; Hendy et al., 2018b) 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, reliability and transparency, it is also essential to share information supporting protein and peptide identification, for example 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 authenticity of endogenous proteins (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., 2025), decontamination with a protocol appropriate for each sample type is recommended prior to protein extraction. Additionally, for example, if 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).

Ultimately, considering that in some cases the volume of evidence obtained in a palaeoproteomic study can be extremely limited, but that a single identification of a species has the potential to substantially rewrite the archaeological narrative, both caution and rigour must be exercised in interpreting the protein data. Not only should a protein identification meet identification standards, there should also be a reasonable expectation of its presence, based on an understanding of the material properties of the sample. For some materials, such as bone, enamel and dentine, there are both detailed proteome-wide studies using modern samples (Alves et al., 2011; Eckhardt et al., 2014; Jágr et al., 2019; Bell et al., 2019), as well as experiments on archaeological samples of known origin (Cappellini et al., 2012; Sawafuji et al., 2017; Ásmundsdóttir et al., 2024). However, for other materials, such as organic residues, there is much less of an understanding of which proteins can be expected to be incorporated, and persist through time. There have been some targeted experiments on beta-lactoglobulin (Ramsøe et al., 2021; Fonseca et al., 2024) and the effects of cooking on protein survival (Barker et al., 2012; Evans et al., 2024; Dekker et al., 2025a), but a generalised understanding remains to be formed. To gain a clearer understanding of the capabilities and limitations of the application of palaeoproteomic methods to these types of materials, we would therefore promote the execution of blind tests and benchmarking studies, as have been done for other archaeological science fields (Barnard et al., 2007; Pestle et al., 2014) and modern proteomics (Van Den Bossche et al., 2021). For metaproteomic analysis, this process is even more essential, as variations in results have been detected depending on the used workflow even in modern metaproteomic studies (Van Den Bossche et al., 2021).

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, for the code see: <https://doi.org/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. Similarly, identification of pathogen or immune-related proteins should have a basis in clinical studies to ensure that conclusions are based on clinically reliable evidence. 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 (Warinner et al., 2014a; Fellows Yates et al., 2021; Ottoni et al., 2021), whereas proteomics has provided more information about past diets (Warinner et al., 2014a; Scott et al., 2021; Wilkin et al., 2021). Additionally, other emerging fields such as palaeometabolomics (Velsko et al., 2017; Brownstein et al., 2020; Barberis et al., 2022; Huber et al., 2023) can be integrated for a comprehensive biomolecular toolkit for analysis of archaeological materials.

Data Sharing and Metadata Availability

The consensus to make data openly available has strengthened and driven development of bioinformatic methods in many fields (e.g., Anagnostou et al., 2015; Lapatas et al., 2015; Celi et al., 2019). 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 (Figure 2). This issue is most prominent for ZooMS studies, where raw data is only available for 26% of all studies that produced data (Figure 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 (Figure 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. (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).

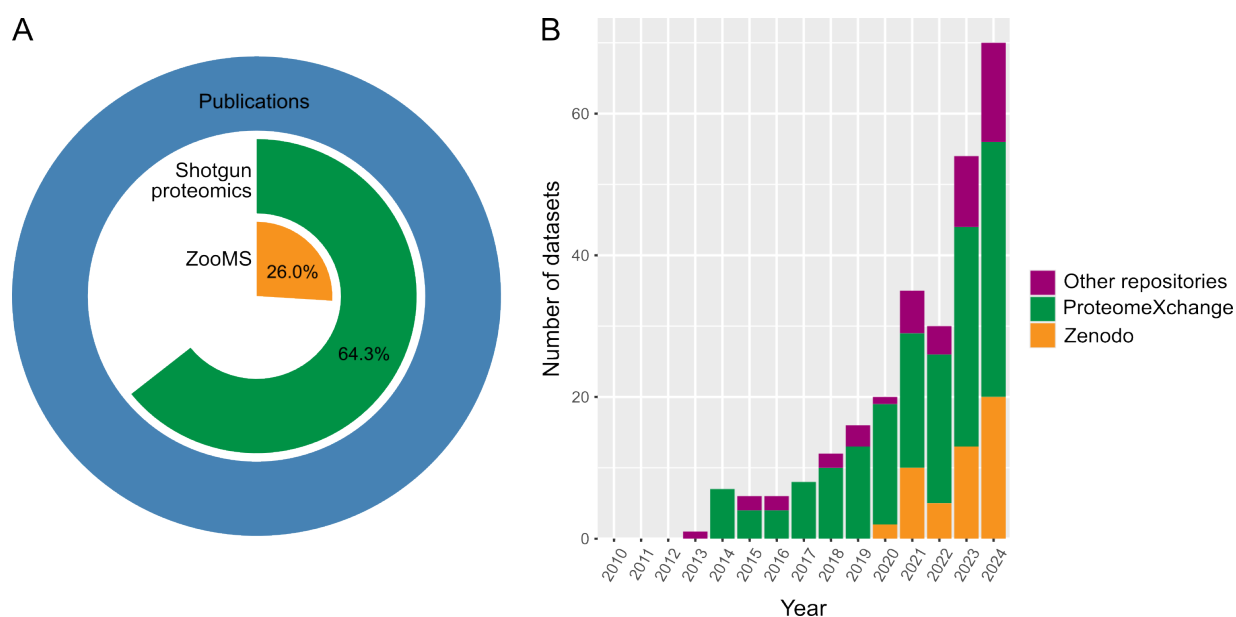


Figure 2 - Data availability in palaeoproteomics. A) Percentage of ZooMS and shotgun palaeoproteomics studies with available datasets. Preprints and papers that did not produce any data were excluded from this calculation. B) Number of palaeoproteomics datasets in online repositories (since 2010). 'Other' repositories include Dryad, Figshare, Github, Mendeley Data, and the OSF Framework. The dataset was compiled by manually going through the list of publications compiled for Figure 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 2009).

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, updated in (Claeys et al., 2023)). 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 of 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, but at a minimum, authors should place accepted versions in institutional repositories to meet public access requirements. However, we are 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. 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. We additionally hope that this paper can serve as a starting point for our collaborators outside of palaeoproteomics (e.g., curators and archaeologists) to understand what standards and approaches our community holds as best practices.

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. Artificial intelligence (AI) systems have revolutionised protein tertiary structure prediction (e.g., Jumper et al., 2021; Tunyasuvunakool et al., 2021), which may impact the future of palaeoproteomics. In palaeoproteomic research, where recovered protein sequences are highly degraded, AI systems may aid researchers in inferring protein structure, thereby improving understanding in evolutionary relationships and functionality (Demarchi et al., 2022), as well as potentially 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 (Chiang & Collins, 2024), 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 and natural 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.

Acknowledgements

We would like to thank the entire PAASTA community for discussions and collaborations which have led to this manuscript. Additionally, we would like to acknowledge that preprint version 4 of this article has been peer-reviewed and recommended by Peer Community in Archaeology (<https://doi.org/10.24072/pci.archaeo.100601>; Daly, 2025) and we would like to thank the reviewers for their valuable suggestions which improved the manuscript.

Funding

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 authors 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.

Conflicts of interest disclosure

We declare that we are in compliance with the PCI rule of having no financial conflicts of interests.

Code and data availability

R code and supporting data have been archived at <https://doi.org/10.5281/zenodo.14899604> (Dekker et al., 2025b)

References

- Abelson PH (1954) Amino acids in fossils. *Science*, 119, 576.
- Alfaro JA, Bohländer P, Dai M, Filius M, Howard CJ, van Kooten XF, Ohayon S, Pomorski A, Schmid S, Aksimentiev A, Anslyn EV, Bedran G, Cao C, Chinappi M, Coyaoud E, Dekker C, Dittmar G, Drachman N, Eelkema R, Goodlett D, Hentz S, Kalathiya U, Kelleher NL, Kelly RT, Kelman Z, Kim SH, Kuster B, Rodriguez-Larrea D, Lindsay S, Maglia G, Marcotte EM, Marino JP, Masselon C, Mayer M, Samaras P, Sarthak K, Sepiashvili L, Stein D, Wanunu M, Wilhelm M, Yin P, Meller A, Joo C (2021) The emerging landscape of single-molecule protein sequencing technologies. *Nature methods*, 18, 604–617. <https://doi.org/10.1038/s41592-021-01143-1>
- Alves RDAM, Demmers JAA, Bezstarosti K, van der Eerden BCJ, Verhaar JAN, Eijken M, van Leeuwen JPTM (2011) Unraveling the human bone microenvironment beyond the classical extracellular matrix proteins: a human bone protein library. *Journal of proteome research*, 10, 4725–4733. <https://doi.org/10.1021/pr200522n>
- Anagnostou P, Capocasa M, Milia N, Sanna E, Battaggia C, Luzi D, Destro Bisol G (2015) When data sharing gets close to 100%: what human paleogenetics can teach the open science movement. *PLoS one*, 10, e0121409. <https://doi.org/10.1371/journal.pone.0121409>
- Ásmundsdóttir RD, Hansen J, Fagernäs Z, Troché G, Olsen JV, Saña Seguí M, Welker F (2024) Early Holocene preservation differences between cortical and trabecular bone proteomes. *Journal of Archaeological Science: Reports*, 57, 104643. <https://doi.org/10.1016/j.jasrep.2024.104643>
- Azémar C, Dufour E, Zazzo A, Wheeler JC, Goepfert N, Marie A, Zirah S (2021) Untangling the fibre ball: Proteomic characterization of South American camelid hair fibres by untargeted multivariate analysis and molecular networking. *Journal of proteomics*, 231, 104040. <https://doi.org/10.1016/j.jprot.2020.104040>
- Bache N, Geyer PE, Bekker-Jensen DB, Hoerning O, Falkenby L, Treit PV, Doll S, Paron I, Müller JB, Meier F, Olsen JV, Vorm O, Mann M (2018) A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Molecular & cellular proteomics: MCP*, 17, 2284–2296. <https://doi.org/10.1074/mcp.TIR118.000853>
- Baker A, Harvey VL, Buckley M (2023) Machine Learning for collagen peptide biomarker determination in the taxonomic identification of archaeological fish remains. *Journal of archaeological science, reports*, 49, 104001. <https://doi.org/10.1016/j.jasrep.2023.104001>
- Barberis E, Manfredi M, Ferraris E, Bianucci R, Marengo E (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, 7208. <https://doi.org/10.3390/molecules27217208>
- Barker A, Venables B, Stevens SM, Seeley KW, Wang P, Wolverton S (2012) An Optimized Approach for Protein Residue Extraction and Identification from Ceramics After Cooking. *Journal of Archaeological Method and Theory*, 19, 407–439. <https://doi.org/10.1007/s10816-011-9120-5>
- Barnard H, Ambrose SH, Beehr DE, Forster, Lanehart RE, Malainey ME, Parr RE, Rider M, Solazzo C, Yohe RM (2007) Mixed results of seven methods for organic residue analysis applied to one vessel with the residue of a known foodstuff. *Journal of archaeological science*, 34, 28–37. <https://doi.org/10.1016/j.jas.2006.03.010>
- Bell PA, Solis N, Kizhakkedathu JN, Matthew I, Overall CM (2019) Proteomic and N-Terminomic TAILS Analyses of Human Alveolar Bone Proteins: Improved Protein Extraction Methodology and

- LysargiNase Digestion Strategies Increase Proteome Coverage and Missing Protein Identification. *Journal of proteome research*, 18, 4167–4179. <https://doi.org/10.1021/acs.jproteome.9b00445>
- Bergström A, Fellows Yates JA, Warinner C (2024) Ancient DNA data hold insights into past organisms and ecosystems - handle them with more care. *Nature*, 636, 296–298. <https://doi.org/10.1038/d41586-024-03993-z>
- Bern M, Kil YJ, Becker C (2012) Byonic: advanced peptide and protein identification software. et al [Current protocols in bioinformatics], Chapter 13, 13.20.1–13.20.14. <https://doi.org/10.1002/0471250953.bi1320s40>
- Brandt LØ, Lillemark MR, Rytter M, Collins MJ, Tøttrup AP (2022) Next Generation Lab puts Denmark's past in the hands of its future. *Antiquity*, 96, 221–228. <https://doi.org/10.15184/aqy.2021.170>
- Brandt LØ, Schmidt AL, Mannering U, Sarret M, Kelstrup CD, Olsen JV, Cappellini E (2014) Species identification of archaeological skin objects from Danish bogs: comparison between mass spectrometry-based peptide sequencing and microscopy-based methods. *PLoS one*, 9, e106875. <https://doi.org/10.1371/journal.pone.0106875>
- Brazil R (2024) Illuminating “the ugly side of science”: fresh incentives for reporting negative results. *Nature*. <https://doi.org/10.1038/d41586-024-01389-7>
- Brown S (2021) Identifying ZooMS Spectra (mammals) using mMass. protocols.io. <https://dx.doi.org/10.17504/protocols.io.bzscp6aw>
- Brown S, Douka K, Collins MJ, Richter KK (2021b) On the standardization of ZooMS nomenclature. *Journal of proteomics*, 235, 104041. <https://doi.org/10.1016/j.jprot.2020.104041>
- Brown S, Hebestreit S, Wang N, Boivin N, Douka K, Korzow Richter K (2020) Zooarchaeology by Mass Spectrometry (ZooMS) for bone material - AmBiC protocol. protocols.io. <https://doi.org/10.17504/protocols.io.bffdjji6>
- Brown S, Kozlikin M, Shunkov M, Derevianko A, Higham T, Douka K, Richter KK (2021a) Examining collagen preservation through glutamine deamidation at Denisova Cave. *Journal of archaeological science*, 133, 105454. <https://doi.org/10.1016/j.jas.2021.105454>
- Brownstein KJ, Tushingham S, Damitio WJ, Nguyen T, Gang DR (2020) An ancient residue metabolomics-based method to distinguish use of closely related plant species in ancient pipes. *Frontiers in molecular biosciences*, 7, 133. <https://doi.org/10.3389/fmolb.2020.00133>
- Buckley M (2015) Ancient collagen reveals evolutionary history of the endemic South American “ungulates.” *Proceedings of the Royal Society B: Biological Sciences*, 282, 20142671. <https://doi.org/10.1098/rspb.2014.2671>
- Buckley M, Collins M, Thomas-Oates J, Wilson JC (2009) Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid communications in mass spectrometry*, 23, 3843–3854. <https://doi.org/10.1002/rcm.4316>
- Buckley M, Fraser S, Herman J, Melton ND, Mulville J, Pálsdóttir AH (2014) Species identification of archaeological marine mammals using collagen fingerprinting. *Journal of archaeological science*, 41, 631–641. <https://doi.org/10.1016/j.jas.2013.08.021>
- Callaway E (2017) Stop hoarding ancient bones, plead archaeologists. *Nature*. <https://doi.org/10.1038/nature.2017.22445>
- Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RAR, Stafford TW, Holen SR, Collins MJ, Orlando L, Willerslev E, Gilbert MTP, Olsen JV (2012) Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *Journal of proteome research*, 11, 917–926. <https://doi.org/10.1021/pr200721u>
- Cappellini E, Welker F, Pandolfi L, Ramos-Madrigal J, Samodova D, Rütther PL, Fotakis AK, Lyon D, Moreno-Mayar JV, Bukhsianidze M, Rakownikow Jersie-Christensen R, Mackie M, Ginolhac A, Ferring R, Tappen M, Palkopoulou E, Dickinson MR, Stafford TW Jr, Chan YL, Götherström A, Nathan SKSS, Heintzman PD, Kapp JD, Kirillova I, Moodley Y, Agusti J, Kahlke R-D, Kiladze G,

- Martínez-Navarro B, Liu S, Sandoval Velasco M, Sinding M-HS, Kelstrup CD, Allentoft ME, Orlando L, Penkman K, Shapiro B, Rook L, Dalén L, Gilbert MTP, Olsen JV, Lordkipanidze D, Willerslev E (2019) Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature*, 574, 103–107. <https://doi.org/10.1038/s41586-019-1555-y>
- Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, Materechera S, Parsons M, Raseroka K, Rodriguez-Lonebear D, Rowe R, Sara R, Walker JD, Anderson J, Hudson M (2020) The CARE principles for indigenous data governance. *Data science journal*, 19. <https://doi.org/10.5334/dsj-2020-043>
- Celi LA, Citi L, Ghassemi M, Pollard TJ (2019) The PLOS ONE collection on machine learning in health and biomedicine: Towards open code and open data. *PLoS one*, 14, e0210232. <https://doi.org/10.1371/journal.pone.0210232>
- Chambers CD, Tzavella L (2022) The past, present and future of Registered Reports. *Nature human behaviour*, 6, 29–42. <https://doi.org/10.1038/s41562-021-01193-7>
- Chen F, Welker F, Shen C-C, Bailey SE, Bergmann I, Davis S, Xia H, Wang H, Fischer R, Freidline SE, Yu T-L, Skinner MM, Stelzer S, Dong G, Fu Q, Dong G, Wang J, Zhang D, Hublin J-J (2019) A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature*, 569, 409–412. <https://doi.org/10.1038/s41586-019-1139-x>
- Chiang Y, Collins MJ (2024) Orthrus: an AI-powered, cloud-ready, and open-source hybrid approach for metaproteomics. *bioRxiv*, 2024.11.15.623814. <https://doi.org/10.1101/2024.11.15.623814>
- Chi H, Liu C, Yang H, Zeng W-F, Wu L, Zhou W-J, Wang R-M, Niu X-N, Ding Y-H, Zhang Y, Wang Z-W, Chen Z-L, Sun R-X, Liu T, Tan G-M, Dong M-Q, Xu P, Zhang P-H, He S-M (2018) Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nature biotechnology*, 36, 1059–1061. <https://doi.org/10.1038/nbt.4236>
- Claeys T, Van Den Bossche T, Perez-Riverol Y, Gevaert K, Vizcaíno JA, Martens L (2023) lesSDRF is more: maximizing the value of proteomics data through streamlined metadata annotation. *Nature communications*, 14, 6743. <https://doi.org/10.1038/s41467-023-42543-5>
- Cleland TP, Schroeter ER, Feranec RS, Vashishth D (2016) Peptide sequences from the first *Castoroides ohioensis* skull and the utility of old museum collections for palaeoproteomics. *Proceedings of the Royal Society B: Biological Sciences*, 283. <https://doi.org/10.1098/rspb.2016.0593>
- Cleland TP, Schroeter ER, Schweitzer MH (2015) Biologically and diagenetically derived peptide modifications in moa collagens. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20150015. <https://doi.org/10.1098/rspb.2015.0015>
- Codlin MC, Douka K, Richter KK (2022) An application of zooms to identify archaeological avian fauna from Teotihuacan, Mexico. *Journal of archaeological science*, 148, 105692. <https://doi.org/10.1016/j.jas.2022.105692>
- Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26, 1367–1372. <https://doi.org/10.1038/nbt.1511>
- Dai C, Füllgrabe A, Pfeuffer J, Solovyeva EM, Deng J, Moreno P, Kamatchinathan S, Kundu DJ, George N, Fexova S, Grüning B, Föll MC, Griss J, Vaudel M, Audain E, Locard-Paulet M, Turewicz M, Eisenacher M, Uszkoreit J, Van Den Bossche T, Schwämmle V, Webel H, Schulze S, Bouyssie D, Jayaram S, Duggineni VK, Samaras P, Wilhelm M, Choi M, Wang M, Kohlbacher O, Brazma A, Papatheodorou I, Bandeira N, Deutsch EW, Vizcaíno JA, Bai M, Sachsenberg T, Levitsky LI, Perez-Riverol Y (2021) A proteomics sample metadata representation for multiomics integration and big data analysis. *Nature communications*, 12, 5854. <https://doi.org/10.1038/s41467-021-26111-3>
- Daly, K. (2025) Emerging standards, challenges and opportunities within palaeoproteomics. *Peer Community in Archaeology*, 100601. <https://doi.org/10.24072/pci.archaeo.100601>

- Dekker JAA, Hagan R, Collins MJ, Hendy J (2025a) An unclear slate, discrepancies between food input and recovered protein signal from experimental foodcrusts. *bioRxiv*, 2025.02.03.636240. <https://doi.org/10.1101/2025.02.03.636240>
- Dekker JAA, Peters C, Winter RM, Collins M, Dickinson MR, Harvey V, Hill E, Nair BAB, Tsutaya T, Viñas Caron L, Warinner C, Welker F, & Fagernäs Z (2025b) Open science, communication, and collaboration for the future of palaeoproteomics. *Zenodo*. <https://doi.org/10.5281/zenodo.15690650>
- Demarchi B, Stiller J, Greal A, Mackie M, Deng Y, Gilbert T, Clarke J, Legendre LJ, Boano R, Sicheritz-Pontén T, Magee J, Zhang G, Bunce M, Collins MJ, Miller G (2022) Ancient proteins resolve controversy over the identity of *Genyornis* eggshell. *Proceedings of the National Academy of Sciences of the United States of America*, 119, e2109326119. <https://doi.org/10.1073/pnas.2109326119>
- Demeter F, Zanolli C, Westaway KE, Joannes-Boyau R, Durringer P, Morley MW, Welker F, Rütger PL, Skinner MM, McColl H, Gauntz C, Vinner L, Dunn TE, Olsen JV, Sikora M, Ponche J-L, Suzzoni E, Frangeul S, Boesch Q, Antoine P-O, Pan L, Xing S, Zhao J-X, Bailey RM, Boualaphane S, Sichanthongtip P, Sihanam D, Patole-Edoumba E, Aubail F, Crozier F, Bourgon N, Zachwieja A, Luangkhot T, Souksavatdy V, Sayavongkhamdy T, Cappellini E, Bacon A-M, Hublin J-J, Willerslev E, Shackelford L (2022) A Middle Pleistocene Denisovan molar from the Annamite Chain of northern Laos. *Nature communications*, 13, 2557. <https://doi.org/10.1038/s41467-022-29923-z>
- Deutsch EW, Bandeira N, Perez-Riverol Y, Sharma V, Carver JJ, Mendoza L, Kundu DJ, Wang S, Bandla C, Kamatchinathan S, Hewapathirana S, Pullman BS, Wertz J, Sun Z, Kawano S, Okuda S, Watanabe Y, MacLean B, MacCoss MJ, Zhu Y, Ishihama Y, Vizcaíno JA (2023) The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic acids research*, 51, D1539–D1548. <https://doi.org/10.1093/nar/gkac1040>
- Dickinson MR, Scott K, Adams NF, Lister AM, Penkman KEH (2024) Amino acid dating of Pleistocene mammalian enamel from the River Thames terrace sequence: a multi-taxon approach. *Quaternary geochronology*, 82, 101543. <https://doi.org/10.1016/j.quageo.2024.101543>
- Dierickx K, Presslee S, Hagan R, Oueslati T, Harland J, Hendy J, Orton D, Alexander M, Harvey VL (2022) Peptide mass fingerprinting of preserved collagen in archaeological fish bones for the identification of flatfish in European waters. *Royal Society open science*, 9, 220149. <https://doi.org/10.1098/rsos.220149>
- Di Gianvincenzo F, Andersen CK, Filtenborg T, Mackie M, Ernst M, Ramos Madrigal J, Olsen JV, Wadum J, Cappellini E (2023) Proteomic identification of beer brewing products in the ground layer of Danish Golden Age paintings. *Science advances*, 9, eade7686. <https://doi.org/10.1126/sciadv.ade7686>
- Ebsen JA, Haase K, Larsen R, Sommer DVP, Brandt LØ (2019) Identifying archaeological leather – discussing the potential of grain pattern analysis and zooarchaeology by mass spectrometry (ZooMS) through a case study involving medieval shoe parts from Denmark. *Journal of cultural heritage*, 39, 21–31. <https://doi.org/10.1016/j.culher.2019.04.008>
- Eckhardt A, Jäger M, Pataridis S, Mikšík I (2014) Proteomic analysis of human tooth pulp: proteomics of human tooth. *Journal of endodontia*, 40, 1961–1966. <https://doi.org/10.1016/j.joen.2014.07.001>
- Engels I, Burnett A, Robert P, Pironneau C, Abrams G, Bouwmeester R, Van der Plaetsen P, Di Modica K, Otte M, Straus LG, Fischer V, Bray F, Mesuere B, De Groote I, Deforce D, Daled S, Dhaenens M (2025) Classification of collagens via peptide ambiguity, in a paleoproteomic LC-MS/MS-based taxonomic pipeline. *Journal of proteome research*, 24, 1907–1925. <https://doi.org/10.1021/acs.jproteome.4c00962>
- Evans M, Hagan R, Boyd O, Bondetti M, Craig O, Collins MJ, Hendy J (2024) The impact of cooking and burial on proteins: a characterisation of experimental foodcrusts and ceramics. *Royal Society open science*, 11. <https://doi.org/10.1098/rsos.240610>

- Evans Z, Paskulin L, Rahemtulla F, Speller CF (2023) A comparison of minimally-invasive sampling techniques for ZooMS analysis of bone artifacts. *Journal of archaeological science, reports*, 47, 103738. <https://doi.org/10.1016/j.jasrep.2022.103738>
- Fagnäs Z, García-Collado MI, Hendy J, Hofman CA, Speller C, Velsko I, Warinner C (2020) A unified protocol for simultaneous extraction of DNA and proteins from archaeological dental calculus. *Journal of archaeological science*, 118, 105135. <https://doi.org/10.1016/j.jas.2020.105135>
- Fagnäs Z, Troché G, Olsen JV, Welker F (2024) Digging deeper into ancient skeletal proteomes through consecutive digestion with multiple proteases. *Journal of proteomics*, 298, 105143. <https://doi.org/10.1016/j.jprot.2024.105143>
- Fagnäs Z, Villa-Islas V, Troché G, Buylaert J-P, Khujageldiev T, Kurbanov R, Olsen JV, Pedersen MW, Welker F (2025) Cleaning the dead: Optimized decontamination enhances palaeoproteomic analyses of a Pleistocene hominin tooth from Khudji, Tajikistan. *Journal of archaeological science*, 106228. <https://doi.org/10.1016/j.jas.2025.106228>
- Fellows Yates JA, Velsko IM, Aron F, Posth C, Hofman CA, Austin RM, Parker CE, Mann AE, Nägele K, Arthur KW, Arthur JW, Bauer CC, Crevecoeur I, Cupillard C, Curtis MC, Dalén L, Díaz-Zorita Bonilla M, Díez Fernández-Lomana JC, Drucker DG, Escribano Esquivá E, Francken M, Gibbon VE, González Morales MR, Grande Mateu A, Harvati K, Henry AG, Humphrey L, Menéndez M, Mihailović D, Peresani M, Rodríguez Moroder S, Roksandic M, Rougier H, Sázelová S, Stock JT, Straus LG, Svoboda J, Teßmann B, Walker MJ, Power RC, Lewis CM, Sankaranarayanan K, Guschanski K, Wrangham RW, Dewhurst FE, Salazar-García DC, Krause J, Herbig A, Warinner C (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. <https://doi.org/10.1073/pnas.2021655118>
- Fiddyment S, Holsinger B, Ruzzier C, Devine A, Binois A, Albarella U, Fischer R, Nichols E, Curtis A, Cheese E, Teasdale MD, Checkley-Scott C, Milner SJ, Rudy KM, Johnson EJ, Vnouček J, Garrison M, McGrory S, Bradley DG, Collins MJ (2015) Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 15066–15071. <https://doi.org/10.1073/pnas.1512264112>
- Fiddyment S, Teasdale MD, Vnouček J, Lévêque É, Binois A, Collins MJ (2019) So you want to do biocodicology? A field guide to the biological analysis of parchment. *Heritage science*, 7, 1–10. <https://doi.org/10.1186/s40494-019-0278-6>
- Fleskes RE, Bader AC, Tsosie KS, Wagner JK, Claw KG, Garrison NA (2022) Ethical guidance in human paleogenomics: New and ongoing perspectives. *Annual review of genomics and human genetics*, 23, 627–652. <https://doi.org/10.1146/annurev-genom-120621-090239>
- Floyd BM, Marcotte EM (2022) Protein sequencing, one molecule at a time. *Annual review of biophysics*, 51, 181–200. <https://doi.org/10.1146/annurev-biophys-102121-103615>
- Fonseca B, Freeman CL, Collins MJ (2024) The survival of β -lactoglobulin peptides in the archaeological record: Vulnerability vs. Sequence variation. *bioRxiv*, 2024.09.13.612646. <https://doi.org/10.1101/2024.09.13.612646>
- Fotakis AK, Denham SD, Mackie M, Orbegozo MI, Mylopotamitaki D, Gopalakrishnan S, Sicheritz-Pontén T, Olsen JV, Cappellini E, Zhang G, Christophersen A, Gilbert MTP, Vågene ÅJ (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, 20190584. <https://doi.org/10.1098/rstb.2019.0584>
- Giraldo O, Garcia A, Corcho O (2018) A guideline for reporting experimental protocols in life sciences. *PeerJ*, 6, e4795. <https://doi.org/10.7717/peerj.4795>
- Gu M, Buckley M (2018) Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. *BMC bioinformatics*, 19, 241. <https://doi.org/10.1186/s12859-018-2221-3>

- Haghighi Z, Mackie M, Apalnes Ørnhøi A, Ramsøe A, Olstad TM, Armitage SJ, Henshilwood CS, Cappellini E (2024) Palaeoproteomic identification of the original binder and modern contaminants in distemper paints from Uvdal stave church, Norway. *Scientific reports*, 14, 12858. <https://doi.org/10.1038/s41598-024-63455-4>
- Hansen J, Dekker J, Troché G, Fagernäs Z, Olsen JV, Saña Seguí M, Welker F (2024) A comparative study of commercially available, minimally invasive, sampling methods on Early Neolithic humeri analysed via palaeoproteomics. *Journal of archaeological science*, 167, 106002. <https://doi.org/10.1016/j.jas.2024.106002>
- Harvey VL, Daugnora L, Buckley M (2018) Species identification of ancient Lithuanian fish remains using collagen fingerprinting. *Journal of archaeological science*, 98, 102–111. <https://doi.org/10.1016/j.jas.2018.07.006>
- Harvey VL, Keating JN, Buckley M (2021) Phylogenetic analyses of ray-finned fishes (Actinopterygii) using collagen type I protein sequences. *Royal Society Open Science*, 8, 201955. <https://doi.org/10.1098/rsos.201955>
- Haug C, Reumer JWF, Haug JT, Arillo A, Audo D, Azar D, Baranov V, Beutel R, Charbonnier S, Feldmann R, Foth C, Fraaije RHB, Frenzel P, Gašparič R, Greenwalt DE, Harms D, Hyžný M, Jagt JWM, Jagt-Yazykova EA, Jarzembowski E, Kerp H, Kirejtshuk AG, Klug C, Kopylov DS, Kotthoff U, Kriwet J, Kunzmann L, McKellar RC, Nel A, Neumann C, Nützel A, Perrichot V, Pint A, Rauhut O, Schneider JW, Schram FR, Schweigert G, Selden P, Szwedo J, van Bakel BWM, van Eldijk T, Vega FJ, Wang B, Wang Y, Xing L, Reich M (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, 413–429. <https://doi.org/10.1007/s12542-020-00522-x>
- Hendy J (2021) Ancient protein analysis in archaeology. *Science advances*, 7. <https://doi.org/10.1126/sciadv.abb9314>
- Hendy J, Warinner C, Bouwman A, Collins MJ, Fiddyment S, Fischer R, Hagan R, Hofman CA, Holst M, Chaves E, Klaus L, Larson G, Mackie M, McGrath K, Mundorff AZ, Radini A, Rao H, Trachsel C, Velsko IM, Speller CF (2018a) Proteomic evidence of dietary sources in ancient dental calculus. *Proceedings of the Royal Society B: Biological Sciences*, 285. <https://doi.org/10.1098/rspb.2018.0977>
- Hendy J, Welker F, Demarchi B, Speller C, Warinner C, Collins MJ (2018b) A guide to ancient protein studies. *Nature ecology & evolution*, 2, 791–799. <https://doi.org/10.1038/s41559-018-0510-x>
- Hickinbotham S, Fiddyment S, Stinson TL, Collins MJ (2020) How to get your goat: automated identification of species from MALDI-ToF spectra. *Bioinformatics (Oxford, England)*, 36, 3719–3725. <https://doi.org/10.1093/bioinformatics/btaa181>
- Hollemeyer K, Altmeyer W, Heinzle E, Pitra C (2008) Species identification of Oetzi's clothing with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry based on peptide pattern similarities of hair digests. *Rapid communications in mass spectrometry*, 22, 2751–2767. <https://doi.org/10.1002/rcm.3679>
- Horsburgh AK (2024) Concluding Commentary. In: *Critical Perspectives on ancient DNA* (eds Strand D, Källén A, Mulcare C), pp. 167–175. The MIT Press. <https://doi.org/10.7551/mitpress/15190.001.0001>
- Huber B, Hammann S, Loeben CE, Jha DK, Vassão DG, Larsen T, Spengler RN, Fuller DQ, Roberts P, Devière T, Boivin N (2023) Biomolecular characterization of 3500-year-old ancient Egyptian mummification balms from the Valley of the Kings. *Scientific reports*, 13, 12477. <https://doi.org/10.1038/s41598-023-39393-y>
- Jágr M, Ergang P, Pataridis S, Kolrosová M, Bartoš M, Mikšík I (2019) Proteomic analysis of dentin-enamel junction and adjacent protein-containing enamel matrix layer of healthy human molar teeth. *European journal of oral sciences*, 127, 112–121. <https://doi.org/10.1111/eos.12594>

- Janzen A, Richter KK, Mwebi O, Brown S, Onduso V, Gatwiri F, Ndiema E, Katongo M, Goldstein ST, Douka K, Boivin N (2021) Distinguishing African bovids using Zooarchaeology by Mass Spectrometry (ZooMS): New peptide markers and insights into Iron Age economies in Zambia. *PloS one*, 16, e0251061. <https://doi.org/10.1371/journal.pone.0251061>
- Jersie-Christensen RR, Lanigan LT, Lyon D, Mackie M, Belstrøm D, Kelstrup CD, Fotakis AK, Willerslev E, Lynnerup N, Jensen LJ, Cappellini E, Olsen JV (2018) Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. *Nature communications*, 9, 4744. <https://doi.org/10.1038/s41467-018-07148-3>
- Jones ED (2022) *Ancient DNA: The Making of a Celebrity Science*. Yale University Press.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. <https://doi.org/10.1038/s41586-021-03819-2>
- Källén A (2025) *The Trouble with ancient DNA: Telling stories of the past with genomic science*. University of Chicago Press, Chicago.
- Källén A, Mulcare C, Nyblom A, Strand D (2024) Petrous fever: The gap between ideal and actual practice in ancient DNA research. *Current anthropology*, 65, 990–1009. <https://doi.org/10.1086/733328>
- Kirby DP, Buckley M, Promise E, Trauger SA, Holdcraft TR (2013) Identification of collagen-based materials in cultural heritage. *The Analyst*, 138, 4849–4858. <https://doi.org/10.1039/c3an00925d>
- Kirby DP, Manick A, Newman R (2020) Minimally Invasive Sampling of Surface Coatings for Protein Identification by Peptide Mass Fingerprinting: A Case Study with Photographs. *Journal of the American Institute for Conservation*, 59, 235–245. <https://doi.org/10.1080/01971360.2019.1656446>
- Kirby DP, Svoboda M, Mazurek J, Spaabæk LR, Southon J (2023) Characterization of an unusual coating on funerary portraits from Roman Egypt circa 100-300AD. *Heritage science*, 11. <https://doi.org/10.1186/s40494-023-00908-5>
- Knudsen GM, Chalkley RJ (2011) The effect of using an inappropriate protein database for proteomic data analysis. *PloS one*, 6, e20873. <https://doi.org/10.1371/journal.pone.0020873>
- Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature methods*, 14, 513–520. <https://doi.org/10.1038/nmeth.4256>
- Lakens D, Mesquida C, Rasti S, Ditroilo M (2024) The benefits of preregistration and Registered Reports. *Evidence-Based Toxicology*, 2. <https://doi.org/10.1080/2833373x.2024.2376046>
- Lapatas V, Stefanidakis M, Jimenez RC, Via A, Schneider MV (2015) Data integration in biological research: an overview. *Journal of biological research (Thessalonike, Greece)*, 22, 9. <https://doi.org/10.1186/s40709-015-0032-5>
- Le Meillour L, Sinet-Mathiot V, Ásmundsdóttir RD, Hansen J, Mylopotamitaki D, Troché G, Xia H, Herrera Bethencourt J, Ruebens K, Smith GM, Fagernäs Z, Welker F (2024) Increasing sustainability in palaeoproteomics by optimizing digestion times for large-scale archaeological bone analyses. *iScience*, 27, 109432. <https://doi.org/10.1016/j.isci.2024.109432>
- Li H, Joh YS, Kim H, Paek E, Lee S-W, Hwang K-B (2016) Evaluating the effect of database inflation in proteogenomic search on sensitive and reliable peptide identification. *BMC genomics*, 17, 1031. <https://doi.org/10.1186/s12864-016-3327-5>
- Lowenstein JM, Sarich VM, Richardson BJ (1981) Albumin systematics of the extinct mammoth and Tasmanian wolf. *Nature*, 291, 409–411. <https://doi.org/10.1038/291409a0>

- Mackie M, Rüther P, Samodova D, Di Gianvincenzo F, Granzotto C, Lyon D, Pegg DA, Howard H, Harrison L, Jensen LJ, Olsen JV, Cappellini E (2018) Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angewandte Chemie*, 57, 7369–7374. <https://doi.org/10.1002/anie.201713020>
- Mann AE, Fellows Yates JA, Fagernäs Z, Austin RM, Nelson EA, Hofman CA (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*. <https://doi.org/10.1016/j.quaint.2020.11.019>
- Miller GH, Beaumont PB, Jull AJ, Johnson B (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, 149–157. <https://doi.org/10.1098/rstb.1992.0092>
- Nair B, Palomo IR, Markussen B, Wiuf C, Fiddymont S, Collins MJ (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. <https://doi.org/10.24072/pcjournal.230>
- Nel TH, Peters C, Richter KK, Henshilwood C, van Niekerk K, Douka K (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. <https://doi.org/10.1016/j.quascirev.2023.108380>
- Ottoni C, Borić D, Cheronet O, Sparacello V, Dori I, Coppa A, Antonović D, Vujević D, Douglas Price T, Pinhasi R, Cristiani E (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. <https://doi.org/10.1073/pnas.2102116118>
- 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.
- Pálsdóttir AH, Bläuer A, Rannamäe E, Boessenkool S, Hallsson JH (2019) Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *Royal Society open science*, 6, 191059. <https://doi.org/10.1098/rsos.191059>
- Parker R, Taylor A, Peng X, Nicastrì A, Zerweck J, Reimer U, Wenschuh H, Schnatbaum K, Ternette N (2021) The choice of search engine affects sequencing depth and HLA class I allele-specific peptide repertoires. *Molecular & cellular proteomics: MCP*, 20, 100124. <https://doi.org/10.1016/j.mcpro.2021.100124>
- Paterson RS, Madupe PP, Cappellini E (2024) Paleoproteomics sheds light on million-year-old fossils. *Nature reviews. Molecular cell biology*, 26, 1–2. <https://doi.org/10.1038/s41580-024-00803-2>
- Patramanis I, Ramos-Madrugal J, Cappellini E, Racimo F (2023) PaleoProPhyler: a reproducible pipeline for phylogenetic inference using ancient proteins. *Peer Community Journal*, 3. <https://doi.org/10.24072/pcjournal.344>
- Penkman KEH, Preece RC, Bridgland DR, Keen DH, Meijer T, Parfitt SA, White TS, Collins MJ (2011) A chronological framework for the British Quaternary based on Bithynia opercula. *Nature*, 476, 446–449. <https://doi.org/10.1038/nature10305>
- Perez-Riverol Y, Bittremieux W, Noble WS, Martens L, Bilbao A, Lazear MR, Grüning B, Katz DS, MacCoss MJ, Dai C, Eng JK, Bouwmeester R, Shortreed MR, Audain E, Sachsenberg T, Van Goey J, Wallmann G, Wen B, Käll L, Fondrie WE (2025) Open-source and FAIR Research Software for Proteomics. *Journal of Proteome Research*, 24, 2222–2234. <https://doi.org/10.1021/acs.jproteome.4c01079>

- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20, 3551–3567. [https://doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2)
- Pestle WJ, Crowley BE, Weirauch MT (2014) Quantifying inter-laboratory variability in stable isotope analysis of ancient skeletal remains. *PloS one*, 9, e102844. <https://doi.org/10.1371/journal.pone.0102844>
- Peters C, Richter KK, Manne T, Dortch J, Paterson A, Travouillon K, Louys J, Price GJ, Petraglia M, Crowther A, Boivin N (2021) Species identification of Australian marsupials using collagen fingerprinting. *Royal Society open science*, 8, 211229. <https://doi.org/10.1098/rsos.211229>
- Peters C, Wang Y, Vakil V, Cramb J, Dortch J, Hocknull S, Lawrence R, Manne T, Monks C, Rössner GE, Ryan H, Siverson M, Ziegler T, Louys J, Price GJ, Boivin N, Collins MJ (2023) Bone collagen from subtropical Australia is preserved for more than 50,000 years. *Communications Earth & Environment*, 4, 1–8. <https://doi.org/10.1038/s43247-023-01114-8>
- Poujois P, Le Meillour L (2023) Palaeoproteomics protocol - arid environment samples. *protocols.io*. <https://doi.org/10.17504/protocols.io.yxmvmn1kog3p/v1>
- Presslee S, Penkman K, Fischer R, Richards-Slidel E, Southon J, Hospitaleche CA, Collins M, MacPhee R (2021) Assessment of different screening methods for selecting palaeontological bone samples for peptide sequencing. *Journal of proteomics*, 230, 103986. <https://doi.org/10.1016/j.jprot.2020.103986>
- Quye A, Strlič M (2019) Ethical sampling guidance. London: The Institute of Conservation (ICON).
- Ramsøe A, Crispin M, Mackie M, McGrath K, Fischer R, Demarchi B, Collins MJ, Hendy J, Speller C (2021) Assessing the degradation of ancient milk proteins through site-specific deamidation patterns. *Scientific reports*, 11, 7795. <https://doi.org/10.1038/s41598-021-87125-x>
- Ramsøe A, van Heekeren V, Ponce P, Fischer R, Barnes I, Speller C, Collins MJ (2020) DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *Journal of archaeological science*, 115, 105080. <https://doi.org/10.1016/j.jas.2020.105080>
- Rayfield EJ, Theodor JM, Polly PD (2020) Fossils from conflict zones and reproducibility of fossil-based scientific data.
- R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Research Data Alliance International Indigenous Data Sovereignty Interest Group (2019) CARE Principles. GIDA-global.org.
- Richter KK, Codlin M, Seabrook M, Warinner C (2022) A primer for ZooMS applications in archaeology. *Proceedings of the National Academy of Sciences*, 119, e2109323119. <https://doi.org/10.1073/pnas.2109323119>
- Rodriguez Palomo I, Nair B, Chiang Y, Dekker J, Dartigues B, Mackie M, Evans M, Macleod R, Olsen JV, Collins MJ (2024) Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *Peer community journal*, 4, 2023.12.15.571577. <https://doi.org/10.24072/pcjournal.491>
- Rusu I, Paica I, Vulpoi A, Radu C, Mircea C, Dobrinescu C, Bodolică V, Kelemen B (2019) Dual DNA-protein extraction from human archeological remains. *Archaeological and anthropological sciences*, 11, 3299–3307. <https://doi.org/10.1007/s12520-018-0760-1>
- Rüther PL, Husic IM, Bangsgaard P, Gregersen KM, Pantmann P, Carvalho M, Godinho RM, Friedl L, Cascalheira J, Taurozzi AJ, Jørkov MLS, Benedetti MM, Haws J, Bicho N, Welker F, Cappellini E, Olsen JV (2022) SPIN enables high throughput species identification of archaeological bone by proteomics. *Nature communications*, 13, 2458. <https://doi.org/10.1038/s41467-022-30097-x>
- Saitta ET, Vinther J, Crisp MK, Abbott GD, Wheeler L, Presslee S, Kaye TG, Bull I, Fletcher I, Chen X, Vidal D, Sanguino F, Buscalioni ÁD, Calvo J, Sereno PC, Baumgart SL, Pittman M, Collins MJ, Sakalauskaite J, Mackie M, Bello FD, Dickinson MR, Stevenson MA, Donohoe P, Heck PR,

- Demarchi B, Penkman KEH (2024) Non-avian dinosaur eggshell calcite can contain ancient, endogenous amino acids. *Geochimica et cosmochimica acta*, 365, 1–20. <https://doi.org/10.1016/j.gca.2023.11.016>
- Santucci VL, Newman PB, Taff BD (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, 239–248. <https://pure.psu.edu/en/publications/toward-a-conceptual-framework-for-assessing-the-human-dimensions->
- Sawafuji R, Cappellini E, Nagaoka T, Fotakis AK, Jersie-Christensen RR, Olsen JV, Hirata K, Ueda S (2017) Proteomic profiling of archaeological human bone. *Royal Society open science*, 4, 161004. <https://doi.org/10.1098/rsos.161004>
- Schroeter ER, Cleland TP (2016) Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid communications in mass spectrometry*, 30, 251–255. <https://doi.org/10.1002/rcm.7445>
- Schroeter ER, DeHart CJ, Cleland TP, Zheng W, Thomas PM, Kelleher NL, Bern M, Schweitzer MH (2017) Expansion for the *Brachylophosaurus canadensis* Collagen I Sequence and Additional Evidence of the Preservation of Cretaceous Protein. *Journal of proteome research*, 16, 920–932. <https://doi.org/10.1021/acs.jproteome.6b00873>
- Scott A, Power RC, Altmann-Wendling V, Artzy M, Martin MAS, Eisenmann S, Hagan R, Salazar-García DC, Salmon Y, Yegorov D, Milevski I, Finkelstein I, Stockhammer PW, Warinner C (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. <https://doi.org/10.1073/pnas.2014956117>
- Scott A, Reinhold S, Hermes T, Kalmykov AA, Belinskiy A, Buzhilova A, Berezina N, Kantorovich AR, Maslov VE, Guliyev F, Lyonnet B, Gasimov P, Jalilov B, Eminli J, Iskandarov E, Hammer E, Nugent SE, Hagan R, Majander K, Onkamo P, Nordqvist K, Shishlina N, Kaverzneva E, Korolev AI, Khokhlov AA, Smolyaninov RV, Sharapova SV, Krause R, Karapetian M, Stolarczyk E, Krause J, Hansen S, Haak W, Warinner C (2022) Emergence and intensification of dairying in the Caucasus and Eurasian steppes. *Nature ecology & evolution*, 6, 813–822. <https://doi.org/10.1038/s41559-022-01701-6>
- Scott A, Warinner C (2020) Ancient Proteins Extraction Protocol. *protocols.io*. <https://doi.org/10.17504/protocols.io.7vwhn7e>
- Sinet-Mathiot V, Rendu W, Steele TE, Spasov R, Madelaine S, Renou S, Soulier M-C, Martisius NL, Aldeias V, Endarova E, Goldberg P, McPherron SJP, Rezek Z, Sandgathe D, Sirakov N, Sirakova S, Soressi M, Tsanova T, Turq A, Hublin J-J, Welker F, Smith GM (2023) Identifying the unidentified fauna enhances insights into hominin subsistence strategies during the Middle to Upper Palaeolithic transition. *Archaeological and anthropological sciences*, 15, 1–18. <https://doi.org/10.1007/s12520-023-01830-4>
- Sinet-Mathiot V, Smith GM, Romandini M, Wilcke A, Peresani M, Hublin J-J, Welker F (2019) Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy). *Scientific reports*, 9, 12350. <https://doi.org/10.1038/s41598-019-48706-z>
- van der Sluis LG, McGrath K, Thil F, Cersoy S, Pétilion J-M, Zazzo A (2023) Identification and tentative removal of collagen glue in Palaeolithic worked bone objects: implications for ZooMS and radiocarbon dating. *Scientific reports*, 13, 22119. <https://doi.org/10.1038/s41598-023-49242-7>
- Society of Vertebrate Paleontology (2025) Member Bylaw on Ethics Statement. Society of Vertebrate Paleontology. <https://vertpaleo.org/member-bylaw-on-ethics-statement/>. Accessed 2025-06-25.
- Soderberg CK, Errington TM, Schiavone SR, Bottesini J, Thorn FS, Vazire S, Esterling KM, Nosek BA (2021) Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature human behaviour*, 5, 990–997. <https://doi.org/10.1038/s41562-021-01142-4>

- Solazzo C, Phipps E (2022) Chasing the elusive viscacha in Precolumbian textiles at the intersection of art and science. *Journal of archaeological science*, 140, 105575. <https://doi.org/10.1016/j.jas.2022.105575>
- Solazzo C, Wilson J, Dyer JM, Clerens S, Plowman JE, von Holstein I, Walton Rogers P, Peacock EE, Collins MJ (2014) Modeling deamidation in sheep α -keratin peptides and application to archeological wool textiles. *Analytical chemistry*, 86, 567–575. <https://doi.org/10.1021/ac4026362>
- Solntsev SK, Shortreed MR, Frey BL, Smith LM (2018) Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of proteome research*, 17, 1844–1851. <https://doi.org/10.1021/acs.jproteome.7b00873>
- Somel M, Altınışık NE, Özer F, Ávila-Arcos MC (2021) Collaborate equitably in ancient DNA research and beyond. *Nature*, 600, 37. <https://doi.org/10.1038/d41586-021-03541-z>
- Speller C, van den Hurk Y, Charpentier A, Rodrigues A, Gardeisen A, Wilkens B, McGrath K, Rowsell K, Spindler L, Collins M, Hofreiter M (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. <https://doi.org/10.1098/rstb.2015.0332>
- Standvoss K, Kazezian V, Lewke BR, Bastian K, Chidambaram S, Arafat S, Alsharif U, Herrera-Melendez A, Knipper A-D, Seco BMS, Soto NN, Rakitzis O, Steinecker I, van Kronenberg Till P, Zarebidaki F, Abbasi P, Weissgerber TL (2024) Shortcut citations in the methods section: Frequency, problems, and strategies for responsible reuse. *PLoS biology*, 22, e3002562. <https://doi.org/10.1371/journal.pbio.3002562>
- Stewart NA, Gerlach RF, Gowland RL, Gron KJ, Montgomery J (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, 13649–13654. <https://doi.org/10.1073/pnas.1714926115>
- Sticker A, Martens L, Clement L (2017) Mass spectrometrists should search for all peptides, but assess only the ones they care about. *Nature methods*, 14, 643–644. <https://doi.org/10.1038/nmeth.4338>
- Strohalm M, Kavan D, Novák P, Volný M, Havlíček V (2010) mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Analytical chemistry*, 82, 4648–4651. <https://doi.org/10.1021/ac100818g>
- Taylor GK, Goodlett DR (2005) Rules governing protein identification by mass spectrometry. *Rapid communications in mass spectrometry*, 19, 3420. <https://doi.org/10.1002/rcm.2225>
- Teasdale MD, Fiddymont S, Vnouček J, Mattiangeli V, Speller C, Binois A, Carver M, Dand C, Newfield TP, Webb CC, Bradley DG, Collins MJ (2017) The York Gospels: a 1000-year biological palimpsest. *Royal Society open science*, 4, 170988. <https://doi.org/10.1098/rsos.170988>
- Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D (2021) Highly accurate protein structure prediction for the human proteome. *Nature*, 596, 590–596. <https://doi.org/10.1038/s41586-021-03828-1>
- University of York (2025) Palaeoproteomics and ZooMS. University of York. <https://www.york.ac.uk/archaeology/research/research-centres/bioarch/research-themes/palaeoproteomics/>
- Van Den Bossche T, Kunath BJ, Schallert K, Schäpe SS, Abraham PE, Armengaud J, Arntzen MØ, Bassignani A, Benndorf D, Fuchs S, Giannone RJ, Griffin TJ, Hagen LH, Halder R, Henry C, Hettich RL, Heyer R, Jagtap P, Jehmlich N, Jensen M, Juste C, Kleiner M, Langella O, Lehmann T, Leith E, May P, Mesuere B, Miotello G, Peters SL, Pible O, Queiros PT, Reichl U, Renard BY, Schiebenhoefer H, Sczyrba A, Tanca A, Trappe K, Trezzi J-P, Uzzau S, Verschaffelt P, von Bergen

- M, Wilmes P, Wolf M, Martens L, Muth T (2021) Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows. *Nature communications*, 12, 7305. <https://doi.org/10.1038/s41467-021-27542-8>
- Végh EI, Douka K (2024) SpecieScan: Semi-Automated taxonomic identification of bone collagen peptides from MALDI-ToF MS spectra. *Bioinformatics*, 40, btae054. <https://doi.org/10.1093/bioinformatics/btae054>
- Velsko IM, Overmyer KA, Speller C, Klaus L, Collins MJ, Loe L, Frantz LAF, Sankaranarayanan K, Lewis CM Jr, Martinez JBR, Chaves E, Coon JJ, Larson G, Warinner C (2017) The dental calculus metabolome in modern and historic samples. *Metabolomics: Official journal of the Metabolomic Society*, 13, 134. <https://doi.org/10.1007/s11306-017-1270-3>
- Venable JD, Dong M-Q, Wohlschlegel J, Dillin A, Yates JR (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature methods*, 1, 39–45. <https://doi.org/10.1038/nmeth705>
- Viñas-Caron LC, Brandt LØ (2025) Clothomics: a practical guide to understand the opportunities and challenges of omics-based methods in archaeological cloth research. *NPJ heritage science*, 13, 80. <https://doi.org/10.1038/s40494-025-01623-z>
- Viñas-Caron LC, Rodríguez Palomo I, Fazlic N, Vnouček J, Driscoll M, Fiddymment S, Collins MJ (2023) A biological reading of a palimpsest. *iScience*, 26, 106786. <https://doi.org/10.1016/j.isci.2023.106786>
- Warinner C, Hendy J, Speller C, Cappellini E, Fischer R, Trachsel C, Arneborg J, Lynnerup N, Craig OE, Swallow DM, Fotakis A, Christensen RJ, Olsen JV, Liebert A, Montalva N, Fiddymment S, Charlton S, Mackie M, Canci A, Bouwman A, Rühli F, Gilbert MTP, Collins MJ (2014a) Direct evidence of milk consumption from ancient human dental calculus. *Scientific reports*, 4, 7104. <https://doi.org/10.1038/srep07104>
- Warinner C, Korzow Richter K, Collins MJ (2022) Paleoproteomics. *Chemical reviews*, 122, 13401–13446. <https://doi.org/10.1021/acs.chemrev.1c00703>
- Warinner C, Rodrigues JFM, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddymment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-García DC, Eppler E, Seiler R, Hansen LH, Castruita JAS, Barkow-Oesterreicher S, Teoh KY, Kelstrup CD, Olsen JV, Nanni P, Kawai T, Willerslev E, von Mering C, Lewis CM Jr, Collins MJ, Gilbert MTP, Rühli F, Cappellini E (2014b) Pathogens and host immunity in the ancient human oral cavity. *Nature genetics*, 46, 336–344. <https://doi.org/10.1038/ng.2906>
- Wehmiller JF, Hare PE, Kujala GA (1976) Amino acids in fossil corals: racemization (epimerization) reactions and their implications for diagenetic models and geochronological studies. *Geochimica et cosmochimica acta*, 40, 763–776. [https://doi.org/10.1016/0016-7037\(76\)90029-6](https://doi.org/10.1016/0016-7037(76)90029-6)
- Welker F, Collins MJ, Thomas JA, Wadsley M, Brace S, Cappellini E, Turvey ST, Reguero M, Gelfo JN, Kramarz A, Burger J, Thomas-Oates J, Ashford DA, Ashton PD, Rowsell K, Porter DM, Kessler B, Fischer R, Baessmann C, Kaspar S, Olsen JV, Kiley P, Elliott JA, Kelstrup CD, Mullin V, Hofreiter M, Willerslev E, Hublin J-J, Orlando L, Barnes I, MacPhee RDE (2015) Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature*, 522, 81–84. <https://doi.org/10.1038/nature14249>
- Welker F, Hajdinjak M, Talamo S, Jaouen K, Dannemann M, David F, Julien M, Meyer M, Kelso J, Barnes I, Brace S, Kamminga P, Fischer R, Kessler BM, Stewart JR, Pääbo S, Collins MJ, Hublin J-J (2016) Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 11162–11167. <https://doi.org/10.1073/pnas.1605834113>
- Welker F, Ramos-Madrigal J, Gutenbrunner P, Mackie M, Tiwary S, Rakownikow Jersie-Christensen R, Chiva C, Dickinson MR, Kuhlwilm M, de Manuel M, Gelabert P, Martínón-Torres M, Margvelashvili A, Arsuaga JL, Carbonell E, Marques-Bonet T, Penkman K, Sabidó E, Cox J, Olsen

- JV, Lordkipanidze D, Racimo F, Lalueza-Fox C, Bermúdez de Castro JM, Willerslev E, Cappellini E (2020) The dental proteome of *Homo antecessor*. *Nature*, 580, 235–238. <https://doi.org/10.1038/s41586-020-2153-8>
- Wickham H (2009) *Ggplot2: Elegant graphics for data analysis*. Springer, New York, NY. <https://doi.org/10.1007/978-0-387-98141-3>
- Wilkin S, Lanigan LT, Montes N, Sharma M, Avanzi C, Sejdiu D, Majander K, Pfrengle S, Chiang Y, Kunz L, Dittmann A, Rühli F, Singh P, Coll MF, Collins MJ, Taurozzi AJ, Schuenemann VJ (2024) Sequential trypsin and ProAlanase digestions unearth immunological protein biomarkers shrouded by skeletal collagen. *iScience*, 27, 109663. <https://doi.org/10.1016/j.isci.2024.109663>
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilkin S, Ventresca Miller A, Fernandes R, Spengler R, Taylor WT-T, Brown DR, Reich D, Kennett DJ, Culleton BJ, Kunz L, Fortes C, Kitova A, Kuznetsov P, Epimakhov A, Zaibert VF, Outram AK, Kitov E, Khokhlov A, Anthony D, Boivin N (2021) Dairying enabled Early Bronze Age Yamnaya steppe expansions. *Nature*, 598, 629–633. <https://doi.org/10.1038/s41586-021-03798-4>
- Wilson J, van Doorn NL, Collins MJ (2012) Assessing the extent of bone degradation using glutamine deamidation in collagen. *Analytical chemistry*, 84, 9041–9048. <https://doi.org/10.1021/ac301333t>
- Winter RM, de Kock W, Mackie M, Ramsøe M, Desiderà E, Collins M, Guidetti P, Presslee S, Alegre MM, Oueslati T, Muniz AM, Michailidis D, van den Hurk Y, Taurozzi AJ, Çakırlar C (2023) Grouping groupers in the Mediterranean: Ecological baselines revealed by ancient proteins. *Ecology and evolution*, 13, e10625. <https://doi.org/10.1002/ece3.10625>
- Xia H, Zhang D, Wang J, Fagernäs Z, Li T, Li Y, Yao J, Lin D, Troché G, Smith GM, Chen X, Cheng T, Shen X, Han Y, Olsen JV, Shen Z, Pei Z, Hublin J-J, Chen F, Welker F (2024) Middle and Late Pleistocene Denisovan subsistence at Baishiya Karst Cave. *Nature*, 1–6. <https://doi.org/10.1038/s41586-024-07612-9>
- Zhang J, Xin L, Shan B, Chen W, Xie M, Yuen D, Zhang W, Zhang Z, Lajoie GA, Ma B (2012) PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & cellular proteomics: MCP*, 11, M111.010587. <https://doi.org/10.1074/mcp.M111.010587>