

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

Item Type	Article
Authors	Dekker, Joannes Adrianus Antonius;Peters, Carli;Winter, Rachel M.;Collins, Matthew J.;Dickinson, Marc R.;Harvey, Virginia L.;Hill, Eden;Nair, Bharath;Tsutaya, Takumi;Viñas-Caron, Laura C.;Warinner, Christina;Welker, Frido;Fagernäs, Zandra
Citation	Dekker, J. A. A., Peters, C., Winter, R. M., Collins, M. J., Dickinson, M. R., Harvey, V. L., Hill, E., Nair, B., Tsutaya, T., Viñas-Caron, L. C., Warinner, C., Welker, F., & Fagernäs, Z. (2025). Open science, communication, and collaboration for the future of palaeoproteomics. Peer Community Journal, 5, article-number e77. <a href="https://doi.org/10.24072/pcjournal.576">https://doi.org/10.24072/pcjournal.576</a>
DOI	<a href="https://doi.org/10.24072/pcjournal.576">10.24072/pcjournal.576</a>
Publisher	Peer Community In
Journal	Peer Community Journal
Download date	2026-05-11 08:27:12
Item License	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
Link to Item	<a href="http://hdl.handle.net/10034/629564">http://hdl.handle.net/10034/629564</a>

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

## *Authors:*

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

## *Affiliations:*

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

## *Corresponding author:*

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

1 **Abstract**

2

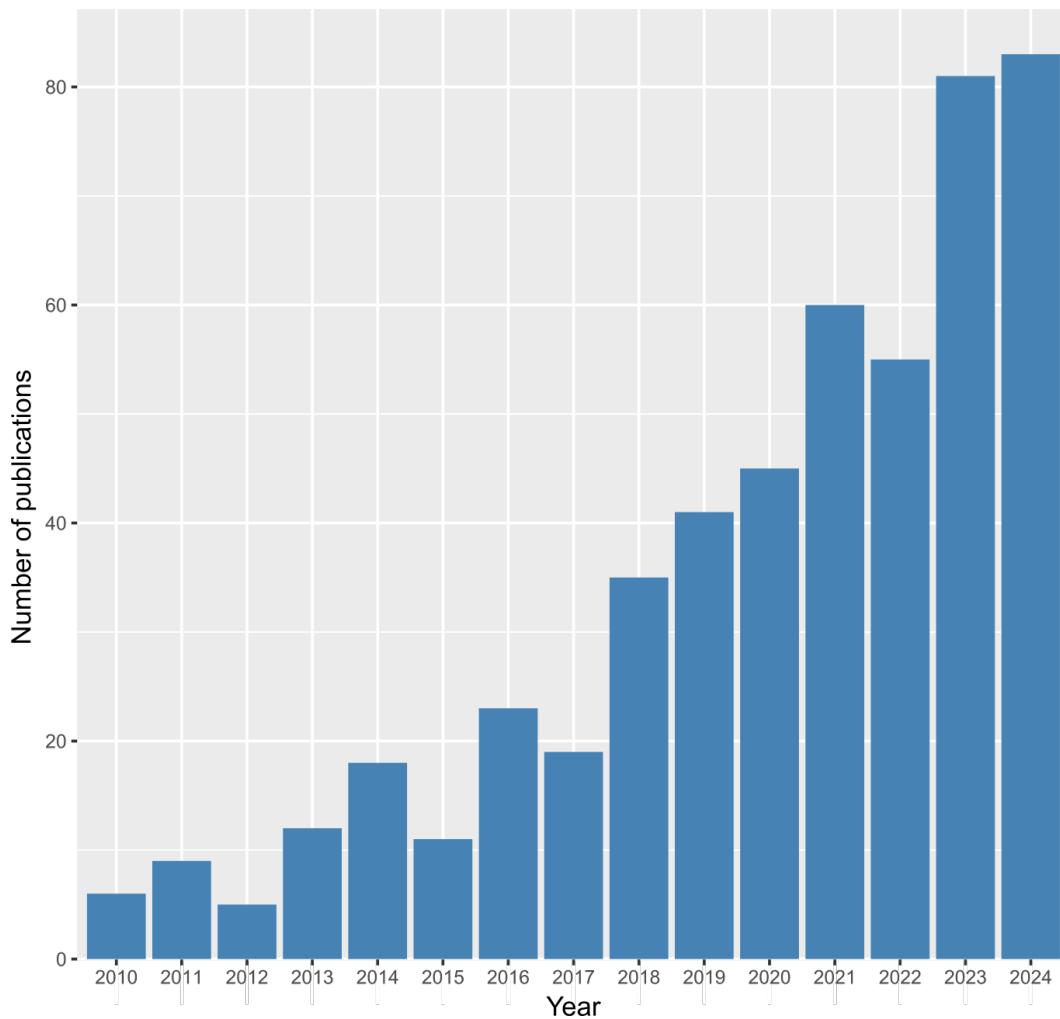
3 Palaeoproteomics focuses on studying the past by analysing proteins preserved in historical,  
4 archaeological, and palaeontological materials. Over the past fifteen years, the number of  
5 publications and research questions in this field has grown substantially, as has the number of  
6 researchers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

17

## 18 Introduction

19

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



27

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

35 *(Wickham, 2016). See also the archived Dataset on Zenodo (10.5281/zenodo.14899604) for a full record*  
36 *of the curated dataset.*

37

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

49 Amino acid sequence variation in proteins between species has led to palaeoproteomic  
50 approaches for taxonomic identification being developed, such as ZooMS (Zooarchaeology by  
51 Mass Spectrometry; Buckley et al., 2009) and SPIN (Species by Proteome INvestigation; R  ther  
52 et al., 2022). These allow for the taxonomic identification of skeletal remains even when no  
53 morphologically identifiable characteristics are present, or when osteomorphology is too similar  
54 between taxa for distinctions. This approach has also been successful in the study of cloth  
55 materials like textiles, fur and leather (Az  mard et al., 2021; Brandt et al., 2014; Hollemeyer et  
56 al., 2008; Solazzo & Phipps, 2022; Vi  nas-Caron & Brandt, 2025), as well as art and heritage  
57 objects such as paintings and manuscripts (Di Gianvincenzo et al., 2023; Fiddyment et al., 2015,  
58 2019; Kirby et al., 2013, 2020, 2023; Teasdale et al., 2017). Furthermore, variation in amino acid  
59 sequences between taxa also allows for phylogenetic placement (Buckley, 2015; Cappellini et al.,  
60 2019; Cleland et al., 2015, 2016; Demarchi et al., 2022; Harvey et al., 2021; Lowenstein et al.,  
61 1981; Schroeter et al., 2017; Welker et al., 2015, 2020), enabling reconstruction of phylogenetic  
62 trees and hypotheses about evolutionary relationships between extinct and extant species. Long-  
63 standing questions in other research fields, such as zooarchaeology, palaeoanthropology,  
64 historical ecology, or zoology, can thereby be addressed through the study of protein sequence  
65 variation.

66 Finally, identification of post-translational modifications (PTMs) of amino acids can provide  
67 valuable information about the history and preservation of proteins. For example, deamidation of  
68 asparagine and glutamine has been proposed as both a preservation or degradation indicator  
69 and a means of assessing the relative preservation of a detected protein or peptide (Brown,  
70 Kozlikin, et al., 2021; Cleland et al., 2015; Rams  e et al., 2020, 2021; Schroeter & Cleland, 2016;  
71 Wilson et al., 2012), and for more specific purposes, such as investigating manufacturing  
72 differences in parchment (Nair et al., 2023). The extent of photo-oxidative modifications can lead  
73 to inferences about past light exposure of artistic or cultural heritage materials, such as paintings  
74 (Mackie et al., 2018). However, caution should be applied when interpreting deamidation for  
75 authentication purposes, as factors other than age can also influence deamidation rates (Brown,  
76 Kozlikin, et al., 2021). Additionally, amino acid racemisation—the spontaneous interconversion of  
77 L-amino acids to their D-enantiomers over time—provides a valuable chronological tool for age

78 determination. Amino acid geochronology has been successfully applied to various materials,  
79 such as shells, eggshells, coral, and enamel, enabling relative dating in palaeontological and  
80 archaeological contexts (Dickinson et al., 2024; Miller et al., 1992; Penkman et al., 2011;  
81 Wehmiller et al., 1976). Evaluating the extent of protein decomposition, through processes like  
82 racemisation, can be useful for screening potential samples for proteomic analysis (Presslee et  
83 al., 2021), and confirming the presence of endogenous amino acids (Cappellini et al., 2019), a  
84 key consideration for older or more degraded samples at the limits of palaeoproteomic analysis.  
85 More broadly, the survival of stable amino acids and the presence of degradation products, first  
86 explored by Abelson (1954), is evidence for the persistence of protein remnants into deep time  
87 (Saitta et al., 2024).

88 The field of palaeoproteomics is thereby highly diverse in terms of both research questions  
89 and methodological approaches. However, researchers in this field commonly face several  
90 issues. For example, ancient proteomes are inherently complex and often contaminated. Even in  
91 cases where only one protein or peptide is the object of research, proteins stemming from the  
92 burial environment and subsequent storage and handling may be present on the studied material,  
93 or introduced during the extraction process (Fagnäs et al., 2025; Hendy, Welker, et al., 2018).  
94 Ancient proteins are also degraded and damaged, complicating analyses (Hendy, 2021). Finally,  
95 the materials that are studied through palaeoproteomics are often rare and/or of significant  
96 cultural heritage value. Precautions should be taken and efforts made to minimise the damage  
97 that is caused when sampling for palaeoproteomic analyses. Since these issues, and many more,  
98 are shared by this research community, the solutions can also be shared.

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

118 At the core of PAASTA is an aim to make the emerging and fast-developing field of  
119 palaeoproteomics more open, communicative, and collaborative. Here, we walk through the usual  
120 steps of a palaeoproteomics project, and discuss how each step could align with these principles,

121 while also integrating the FAIR (guidelines for making data Findable, Accessible, Interoperable  
122 and Reusable; Wilkinson et al., 2016) and CARE (guidelines for sharing Indigenous Peoples data;  
123 Collective benefit, Authority to control, Responsibility and Ethics; Research Data Alliance  
124 International Indigenous Data Sovereignty Interest Group, 2019) principles. We are aware that  
125 each archaeological science project is unique and all suggestions below are unlikely to be  
126 applicable to all projects. However, by discussing and promoting different options and guidelines,  
127 we can establish a baseline for this community and collectively strengthen the research field step-  
128 by-step.

## 129 Experimental Design

130  
131 The first step in any palaeoproteomics project is setting up an appropriate research design and  
132 sampling strategy that can answer the specific research question asked. Such a strategy should  
133 be thoroughly discussed with all relevant stakeholders and planned prior to sampling. Depending  
134 on the context of the material being studied, ethical sampling guidelines by discipline may already  
135 be in place and available for consultation (e.g. Institute of Conservation Heritage Science Group  
136 (Quye & Strlič, 2019), Paleontological Research Institution’s Special Terms and Conditions for  
137 Technical Analysis/Destructive Sampling). Such documents can provide palaeoproteomics  
138 researchers with initial insight into the perspectives and guidelines of their collaborators. Various  
139 factors should be taken into account, such as preservation differences within samples  
140 (Ásmundsdóttir et al., 2024) or sites (Le Meillour et al., 2024), sample treatment (e.g., legacy  
141 samples or museum specimens; (Haghighi et al., 2024) and manufacturing processes (e.g. liming,  
142 tanning, dyeing; (Ebsen et al., 2019; Nair et al., 2023; Solazzo et al., 2014). Curators have  
143 become increasingly wary about destructive analysis as a result of the “... *tension between the*  
144 *professional obligations of curators to faithfully steward archaeological collections and ensure*  
145 *their ongoing integrity and the “harvesting” desires of specialists intent on industrial-scale aDNA*  
146 *research in the absence of a research question*” (Källén et al., 2024). To avoid a breakdown in  
147 relationships between communities, archaeologists and museums, we must ensure that  
148 destructive sampling is minimised, thoroughly discussed and understood by all parties in advance,  
149 and that additional research opportunities have been carefully considered. Wherever possible,  
150 the use of ‘self-sampled’ material—fragments that have already detached from the object or  
151 material—should be prioritised to avoid damage to the main mass. Conversations supported by  
152 visual aids (e.g., Kirby et al., 2020) demonstrating the minimal sample sizes and minimally  
153 destructive approaches used in palaeoproteomics (Fiddymment et al., 2015) can help illustrate how  
154 the field is striving to balance information gain with the preservation of cultural and natural heritage  
155 materials. There is no single right answer for this tradeoff between preservation and obtaining  
156 new information through (potentially) destructive analysis, rather it will be extremely case-specific,  
157 depending on the amount of material available, the research question and the risk of an  
158 unsuccessful analysis. The core of conversations with curators and other stakeholders should  
159 therefore be to obtain ‘informed consent’, with all parties subscribing to the aim of the analysis  
160 and being aware of the risks involved. Given that the methods for extracting aDNA, proteins, and  
161 lipids target different physicochemical characteristics of each material, it should be the intention

162 to sample in a way that maximizes the potential to generate complementary datasets from the  
163 same sub-samples (e.g., for archaeogenetic analysis, isotopic analysis, radiocarbon dating).  
164

165 Palaeoproteomic analyses are destructive to varying degrees, as part of a specimen is  
166 irreversibly removed, but the degree of destructiveness varies depending on the chosen  
167 extraction protocol. Minimally destructive protocols have also been developed, but vary in their  
168 success levels (Z. Evans et al., 2023; Hansen et al., 2024). Open and collaborative  
169 communication with those that provide us with material to sample (communities, archaeologists,  
170 conservators, collection managers, etc.), addressing the risks of the study and the extent of  
171 sampling destructiveness (e.g., the amount of material required and what will remain), is essential  
172 for balancing the needs of palaeoproteomics research with the conservation of objects or remains  
173 (Fiddymment et al., 2015; Fleskes et al., 2022; Pálsdóttir et al., 2019). This will also aid in  
174 understanding the conservation history of the sample, which may influence which analyses can  
175 be conducted, such as the introduction of modern contaminants or treatment with animal-based  
176 glues (Haghighi et al., 2024; van der Sluis et al., 2023). During the sampling process, it is also  
177 recommended to take high-quality photos before and after sampling of the specimen; additionally,  
178 institutions curating human remains may require 3D or CT scanning of remains prior to destructive  
179 analysis, which adds to the cost. Documentation prior to and after sampling will both aid in  
180 identifying visual indicators that distinguish successful/unsuccessful samples and promote  
181 transparency regarding the traces left by sampling methods on ancient specimens.  
182

183 Collaboration across different research specialties is both essential and highly beneficial  
184 for the success of a research project. When working with material that originates directly from an  
185 archaeological site, it is vital to work collaboratively with the site specialists, excavation teams,  
186 and any other specialists involved. They have the site-specific knowledge about stratigraphy,  
187 taphonomy, and chronology of the archaeological contexts, all of which should be taken into  
188 account when selecting which material is suitable to sample for palaeoproteomics research. The  
189 value of any archaeological object for our understanding of the past is highly dependent on the  
190 archaeological context it was found in. Engaging with archaeologists, curators or other  
191 stakeholders from the earliest stages of experiment design can be immensely helpful in ensuring  
192 that the selected samples match the research questions and that the research questions  
193 themselves will meaningfully contribute to advancing our understanding of the past.  
194 Palaeoproteomic research offers the opportunity to open up new lines of questioning and address  
195 research questions that other methodologies are not well suited for, but its full value can only be  
196 achieved through integration in theoretical frameworks including data obtained from other  
197 methodologies. For example, in the case of ZooMS and SPIN, active collaboration with  
198 zooarchaeologists that work with the faunal material is crucial to the development of a suitable  
199 sampling strategy that not only takes into account the palaeoproteomic perspective, but that also  
200 helps address specific research questions stemming from zooarchaeology (Hansen et al., 2024;  
201 Sinet-Mathiot et al., 2019, 2023). Additionally, an understanding of the faunal spectrum is  
202 essential for maximising the specificity of the taxonomic identifications made with methods like  
203 ZooMS and SPIN. Whilst certain taxa may share the same peptide sequences, they can often be  
204 distinguished based on their temporal and geographic distribution. Further, closer collaboration  
205 with specialists in other fields in biomolecular archaeology (e.g., isotope ratio mass spectrometry,

206 archaeogenomics, and organic residue analysis) can lead to the development of simultaneous or  
207 sequential extraction protocols that minimise the need for duplicated destructive sampling  
208 (Fagernäs et al., 2020; Rusu et al., 2019). Where possible, any remaining material following  
209 analyses should be carefully returned to collections to support future research opportunities and  
210 reduce the need for additional sampling. Finally, collaboration with bioinformaticians and  
211 statisticians can help enhance data analysis and interpretation, as well as ensure that the  
212 research design has the power to answer the research question. While a single sample may be  
213 enough for taxonomic identification of a specimen, other applications such as method  
214 development and comparisons between populations or archaeological sites need more samples,  
215 and it is essential to ensure that a statistically robust experimental design is set up prior to  
216 sampling. We also want to emphasize the necessity of equality between fields in a  
217 multidisciplinary study. All methods, including biomolecular techniques such as  
218 palaeoproteomics, but also macro- or microscopic methodologies, have their own limitations and  
219 advantages. The added value of multidisciplinary lies in the integration of the contributions from  
220 different analyses in a single framework.

221

222 The rights of Indigenous peoples and the colonial nature of resource-intensive ancient  
223 biomolecular research have long been debated in archaeology and archaeological science, and  
224 most recently heavily discussed in the field of archaeogenetics (Fleskes et al., 2022; Källén, 2025;  
225 Källén et al., 2024; Somel et al., 2021). As an emerging field, palaeoproteomics can learn from  
226 the discussion on the ethics of archaeological research and has the potential to democratise  
227 ancient biomolecular research by promoting communication and collaboration with stakeholders  
228 and researchers in the country of origin of the samples (Paterson et al., 2024). Participation in  
229 palaeoproteomic research can also be broadened through citizen science projects (Brandt et al.,  
230 2022). Since proteins can be analysed from materials that are significantly older than those used  
231 for aDNA studies (Cappellini et al., 2019; Demarchi et al., 2022; Harvey et al., 2018; Welker et  
232 al., 2020), the ethical dimensions of palaeontological research, which have not been much  
233 discussed in aDNA research, may become a challenge for palaeoproteomics. For example,  
234 palaeontological specimens are sometimes commercially traded or privately owned. Although  
235 more rarely the case in archaeology, similar ethical concerns arise in the study of privately owned  
236 or commercially traded art objects, such as paintings, where issues of provenance, access, and  
237 research use must also be carefully considered. The research use of such specimens is debated  
238 in palaeontology (Haug et al., 2020; Rayfield et al., personal communication, 2020; Santucci et  
239 al., 2016), and an open interdisciplinary discussion among researchers within palaeoproteomics  
240 is essential to establish clear ethical guidelines that balance scientific advancement with respect  
241 for the cultural, legal and ownership complexities surrounding these materials. Existing  
242 frameworks, such as the Society for Vertebrate Paleontology's ethical guidelines regarding the  
243 use of commercial specimens (Society of Vertebrate Paleontology, 2025), may provide useful  
244 models for developing such standards.

245

## 246 Protein Extraction and Data Acquisition

247

248 Protein extraction protocols are continuously being developed to increase the amount of high-  
249 quality data that is obtained from each sample. Sharing laboratory protocols is essential for  
250 reproducibility of studies and for advancing the field as a whole. Methods sections in traditional  
251 journal publications are, however, often limited in the amount of detail that can be included.  
252 Recreating a protocol based on methods reported within an article can therefore be highly  
253 challenging. Furthermore, citations to methods in previous publications often lead to issues with  
254 finding the original protocol, including broken citation chains and publications being paywalled  
255 (Standvoss et al., 2024). Sharing detailed stepwise protocols on open platforms, such as  
256 protocols.io, ensures that the method can be accurately adapted by other researchers, while also  
257 ensuring that the original publication is cited, and that different adaptations of the same protocol  
258 can be traced using a version-tracking system (e.g. Brown et al., 2020; Poujois & Le Meillour,  
259 2023; Scott & Warinner, 2020). For methods sections in journal articles, guidelines have been  
260 created for what should be included (Giraldo et al., 2018), containing details ranging from the  
261 protocol version that is used, to the precise equipment, consumables and reagents, and exact  
262 procedures used at each step. This level of detailed method sharing ensures that for each  
263 destructive sampling of archaeological specimens, the most appropriate protocol can be selected  
264 and applied, irrespective of which research group is analysing the specimen.

265

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

279

280 Mass spectrometry methods are also constantly developing and improving; however,  
281 researchers in the field of palaeoproteomics are rarely involved in such methodological  
282 improvements to instrumentation, in part because many palaeoproteomics groups rely on core  
283 facilities where they may have limited input or control over instrument operations. Closer  
284 collaboration with researchers in the broader fields of protein mass spectrometry and proteomics  
285 may lead to significant steps forward in palaeoproteomics due to the faster adoption of new  
286 methodologies. This would ensure both that new developments are also considering the needs  
287 of palaeoproteomics, as well as that palaeoproteomics researchers are aware of the limitations  
288 of the applicability of new methodology to their research. Moreover, such collaborations might  
289 also lead to the development of methods specifically designed to suit the challenges of

290 palaeoproteomics. For example, SPIN (Rüther et al., 2022) was developed for palaeoproteomics  
291 in parallel to the development of bead-based protein capture and digestion approaches, EvoSep-  
292 based liquid chromatography workflows (Bache et al., 2018), and DIA data acquisition strategies  
293 (Venable et al., 2004). These developments are primarily aimed for clinical proteomics, but are  
294 also suitable for the analysis of ancient proteins. In the context of palaeoproteomics, efforts to  
295 standardise LC-MS/MS workflows, such as the adoption of systems with predefined, reproducible  
296 chromatographic settings, can facilitate direct comparison between datasets and enhance  
297 reproducibility across experiments and laboratories. Although the use of fixed chromatographic  
298 methods limits the degree of optimisation for specific sample types, it promotes consistency,  
299 which is critical for building comparable palaeoproteomic datasets across the research  
300 community.

301  
302 As manuscripts typically focus on successful protocols and analyses, negative results are  
303 often not shared with the scientific community. We recognize that this may be due to pressure to  
304 publish polished high-impact studies, however, avoiding dissemination of negative results is in  
305 the end detrimental to the field as a whole. This may lead to repeated unsuccessful attempts at  
306 using a specific method, or analysing a particular type of sample, as researchers may not be  
307 aware of previous negative outcomes due to a lack of communication. In addition to wasting time  
308 and funding, this also risks a waste of archaeological samples, and may overinflate the image of  
309 how successful certain types of studies are. Often, publications featuring positive results can also  
310 publish associated negative results, such as a failure to assign taxonomy (Demeter et al., 2022),  
311 absence of protein preservation in a certain tissue (Chen et al., 2019) or at certain sites (Peters  
312 et al., 2023), or unexpectedly poor performance of a protease (Fagernäs et al., 2024). Mentioning  
313 negative results in the main text (as opposed to only being included in the supplementary  
314 information) will ensure that they reach the scientific community. Journals are increasingly willing  
315 to publish high-quality research that led to negative results, and specialised journals have also  
316 been founded purely for the purpose of reporting negative results (Brazil, 2024). Negative results  
317 can also be shared at conferences, where they will reach a relevant audience, and associated  
318 conference abstracts may be published and become citable resources (e.g. Paleoanthropology  
319 Society & European Society for the study of Human Evolution, 2024), although it should be noted  
320 that the amount of detail that can be shared is very limited, and as such, evaluating negative  
321 results is challenging. Additionally, preprints can be used to share reports on negative results,  
322 and empty or otherwise failed raw data files can be shared in data repositories. Finally, as  
323 colleagues, reviewers and editors, we can collectively avoid stigmatising non-significant test  
324 outcomes or negative results.

325  
326 Alternatively, the recent development of pre-registering publications, known as Registered  
327 Reports, aims to circumvent the issues associated with publishing negative results (Chambers &  
328 Tzavella, 2022). The details of what needs to be included in a Registered Report differ between  
329 journals, but the core concept is to submit a research plan including hypotheses, methodologies  
330 and planned analyses (Soderberg et al., 2021), which is subjected to peer review and can be  
331 either published separately or result in an in-principle acceptance of a later publication by a  
332 journal. This form of publication moves the peer-review to the start of the publication process

333 (Lakens et al., 2024), so that an experiment is judged by the merits of its planning and negative  
334 results will not reduce opportunities for publication.

## 335 Data Analysis and Interpretation

336  
337 With improvements in protein extraction and mass spectrometry technology, and accompanying  
338 increased sizes of palaeoproteomic datasets, there is an increasing necessity of field-specific or  
339 adaptable software and pipelines for data analysis. In the wider proteomics community there is a  
340 call for increasing use of FAIR and open-source software (Perez-Riverol et al., 2024), and these  
341 principles should be applied also to palaeoproteomics analysis tools, to ensure transparency and  
342 replicability. Development, comparisons and benchmarking of software and pipelines can greatly  
343 benefit from a community-wide effort and communication, ensuring wide applicability and  
344 reliability across study materials, research groups and computing systems.

345  
346 The analysis of palaeoproteomic data is fundamentally different between ZooMS peptide  
347 mass fingerprinting and shotgun proteomics experiments. ZooMS taxonomic identifications are  
348 often obtained by manually comparing the observed peaks to a reference list of biomarkers or  
349 reference spectra. Some pre-processing steps of the spectra have become standard, such as  
350 averaging triplicate spectra, automatic peak picking and setting minimum signal-to-noise  
351 thresholds, but the association between an observed peak and a biomarker remains a manual  
352 step in most studies. Consequently, there is potential for a degree of subjectivity and inter-  
353 observer variability. To address this issue, there have been improvements in the development of  
354 automated tools for biomarker identification (Baker et al., 2023; Gu & Buckley, 2018;  
355 Hickinbotham et al., 2020; Végh & Douka, 2024). Such tools are of great interest and utility to the  
356 palaeoproteomics community, if the developed tools are made freely and openly available (e.g.,  
357 <https://github.com/touzet/pampa/>). Reporting and interpretation guidelines have been established  
358 for ZooMS (Brown, 2021) using mMass (Strohalm et al., 2010), which is currently the most popular  
359 tool for analysis of ZooMS data. It is a freely available and user-friendly software tool that easily  
360 allows the processing and display of MS1 spectra, but is no longer supported by its developer  
361 (though it remains accessible on GitHub: <https://github.com/xxao/mMass>). The development of a  
362 new open-access standard tool for visualising and processing ZooMS data would therefore be of  
363 great value to the palaeoproteomics community.

364  
365 Of great importance to ZooMS analysis is the choice of biomarkers with which to compare  
366 the spectra. Since the initial publication of seven collagen peptide biomarkers for a small selection  
367 of taxa (Buckley et al., 2009), the repertoire of these biomarkers has been expanded to an  
368 increasingly wide range of species (Buckley et al., 2014; Speller et al., 2016; Welker et al., 2016;  
369 Xia et al., 2024). An overview of all available biomarkers can be found in a Google Spreadsheet  
370 kept by the University of York (University of York, 2025). This is a live document that is  
371 continuously being updated as new biomarkers are published. A key study in the biomarker  
372 development has been the standardisation of biomarker nomenclature and their explicit coupling  
373 to specific peptides rather than primarily being based on particular *m/z* windows (Brown, Douka,  
374 et al., 2021). This has facilitated the communication, discussion and validation of new biomarkers

375 (Codlin et al., 2022; Dierickx et al., 2022; Janzen et al., 2021; Nel et al., 2023; Peters et al., 2021;  
376 Winter et al., 2023) and sets a standard against which proposed homologous biomarkers between  
377 taxa can be compared. Furthermore, the increasing number of species for which biomarkers are  
378 available has resulted in the rise of community-based and collaborative projects that aim to  
379 enhance accessibility, as well as ensure robustness through independent validation of  
380 biomarkers; something that the PAASTA community aims to play a key role in.

381  
382 In contrast to the predominantly manual data analysis in ZooMS, shotgun proteomics  
383 experiments utilise protein search engines to match observed precursor and fragment ion spectra  
384 to peptide and protein databases. There are a large number of different protein search engines  
385 available, some of which are free to use - e.g., MaxQuant (Cox & Mann, 2008), Fragpipe (Kong  
386 et al., 2017), MetaMorpheus (Solntsev et al., 2018) and pFind (Chi et al., 2018), while others  
387 require a paid license - e.g., Mascot (Perkins et al., 1999), PEAKS (Zhang et al., 2012), and  
388 Byonic (Bern et al., 2012). The number of options in protein search engines can make it  
389 overwhelming to make appropriate selections. Key features that help distinguish between different  
390 programmes are *de novo* sequencing and open searches. The usefulness of such features for  
391 data analysis depends greatly on the nature of the samples and the aim of the analysis. *De novo*  
392 sequencing is often applied in cases where it is known that there is no proteomic reference data  
393 available for all expected target species, and open searches are frequently applied when a high  
394 number of PTMs are expected. Beyond features such as open searches or *de novo* sequencing,  
395 there are more subtle differences in the scoring algorithm used to identify peptides and proteins.  
396 Consequently, there are differences in the number of peptides and proteins that different protein  
397 search engines identify (Rodriguez Palomo et al., 2024), as well as which peptides and proteins  
398 they identify (Parker et al., 2021). It may therefore not always be possible to completely replicate  
399 the same peptide and protein identifications when analysing the same data using a different  
400 protein search engine. This also means that the use of proprietary software for protein analysis  
401 imposes a barrier to the reproducibility of the experiment. Furthermore, specific software may be  
402 more suited for the identification of short peptides, or high numbers of PTMs, and therefore  
403 comparing different search engines before deciding which one to use in the analysis may lead to  
404 improved protein and peptide identifications.

405  
406 Pipelines for the analysis of shotgun proteomics datasets increase reproducibility and  
407 transparency in palaeoproteomics studies. Such pipelines have, for example, been developed for  
408 phylogenetic analyses based on palaeoproteomic data, including both the creation of a set of  
409 reference sequences translated from genomic sequences, as well as subsequent phylogenetic  
410 analysis (Patramanis et al., 2023). Additionally, a pipeline has been written for collagen-based  
411 taxonomic classification through LC-MS/MS data, which is also capable of identifying mixtures of  
412 species in a sample (Engels et al., 2025). Further development of such pipelines, shared openly  
413 and including thorough documentation and tutorials, will help the entire palaeoproteomics  
414 community, as they can be applied to different datasets and be thoroughly tested by different  
415 researchers.

416  
417 Apart from the choice of a particular protein search engine, the choice of a protein  
418 reference database has a large impact on the identification of peptides and proteins. The most

419 straightforward effect is that any proteins not present in the database will not be identified in the  
420 sample. Instead, ions belonging to a protein present in the sample, but not in the protein reference  
421 database, may be erroneously assigned to a different protein or taxon (Knudsen & Chalkley,  
422 2011). Additionally, the relative size of the database, in terms of the number of protein sequences  
423 included, compared to the number of proteins present in the sample, can also impact protein  
424 identification (Rodriguez Palomo et al., 2024). This effect is described elsewhere (Li et al., 2016)  
425 in more detail, but perhaps the most significant implication of this phenomenon is that relatively  
426 small and constrained databases are preferred compared to broad databases (Sticker et al.,  
427 2017). Despite this, the latter are still frequently used especially in the analysis of organic residues  
428 where little can be assumed regarding the composition of the sample. Finally, protein sequences  
429 available in reference databases can be incorrectly annotated (Harvey et al., 2021), which, if not  
430 caught, can cause issues for downstream interpretations. Considering the substantial effect of  
431 the composition of the database on the outcome of the analysis, many researchers create their  
432 own custom targeted databases. Although the use of custom databases has many advantages  
433 for protein analysis, it can complicate matters for those reviewing or interpreting the published  
434 data. If the database itself is not included in the publication it can be challenging to recreate it  
435 and, consequently, to replicate the analysis. Moreover, detailed metadata, including protein  
436 accession numbers, species names, and taxonomic identification numbers, are required to  
437 understand which proteins of which taxa are included in the database and therefore what taxa  
438 can or cannot be excluded as potential sources of the observed proteins. It is therefore  
439 recommended to either share the database or to provide highly detailed metadata for recreation  
440 of the database.

441  
442 Considering the risk to reproducibility that comes with variety in protein analysis software,  
443 pipelines and databases, we reiterate earlier metadata reporting standards (Hendy, Welker, et  
444 al., 2018; Taylor & Goodlett, 2005) in asking that any palaeoproteomic publication should strive  
445 for maximum openness regarding used parameters and at the minimum contain:

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

459  
460 To ensure reproducibility, reliability and transparency, it is also essential to share  
461 information supporting protein and peptide identification, e.g. as part of a ProteomeXchange  
462 project or as supplementary files of a manuscript. For identified proteins, at the very least the

463 accession number from the used database should be shared, to ensure that it is possible to trace  
464 the origin and additional information for the identification (such as protein name and taxonomy).  
465 Additionally, the number of supporting peptides and the peptides themselves should be shared.  
466 For identified peptides, the amino acid sequence and the assigned start and end positions within  
467 the protein are important information to share, as well as the number of supporting spectra, and  
468 any PTMs and their location. Sharing this information allows reuse of data without the need to re-  
469 generate protein and peptide identifications, as well as independent analysis of the results.

470  
471 For ancient proteins, assessing their authenticity is essential, as they are often low-  
472 abundance and found in a mixture of proteins originating from a variety of sources (Hendy,  
473 Welker, et al., 2018). Reporting deamidation values for proteins or proteomes of interest, and  
474 comparing them with laboratory blanks or more recent samples, can aid in assessing the  
475 authenticity of endogenous proteins (Ramsøe et al., 2020). Additionally, processing and analysing  
476 laboratory blanks alongside samples will ensure that contamination from the laboratory or  
477 reagents are not interpreted as endogenous to the samples. As contamination has been shown  
478 to impact the reconstructed ancient proteome (Fagernäs et al., 2025), decontamination with a  
479 protocol appropriate for each sample type is recommended prior to protein extraction.  
480 Additionally, for example, if dietary items are targeted in a study, analysing samples from the  
481 surrounding sediment or other tissues may ensure that environmental taxa are not mistakenly  
482 assumed to stem from the sample of interest (Mann et al., 2020).

483  
484 Ultimately, considering that in some cases the volume of evidence obtained in a  
485 palaeoproteomic study can be extremely limited, but that a single identification of a species has  
486 the potential to substantially rewrite the archaeological narrative, both caution and rigour must be  
487 exercised in interpreting the protein data. Not only should a protein identification meet  
488 identification standards, there should also be a reasonable expectation of its presence, based on  
489 an understanding of the material properties of the sample. For some materials, such as bone,  
490 enamel and dentine, there are both detailed proteome-wide studies using modern samples (Alves  
491 et al., 2011; Bell et al., 2019; Eckhardt et al., 2014; Jágr et al., 2019), as well as experiments on  
492 archaeological samples of known origin (Ásmundsdóttir et al., 2024; Cappellini et al., 2012;  
493 Sawafuji et al., 2017). However, for other materials, such as organic residues, there is much less  
494 of an understanding of which proteins can be expected to be incorporated, and persist through  
495 time. There have been some targeted experiments on beta-lactoglobulin (Fonseca et al., 2024;  
496 Ramsøe et al., 2021) and the effects of cooking on protein survival (Barker et al., 2012; Dekker  
497 et al., 2025; M. Evans et al., 2024), but a generalised understanding remains to be formed. To  
498 gain a clearer understanding of the capabilities and limitations of the application of  
499 palaeoproteomic methods to these types of materials, we would therefore promote the execution  
500 of blind tests and benchmarking studies, as have been done for other archaeological science  
501 fields (Barnard et al., 2007; Pestle et al., 2014) and modern proteomics (Van Den Bossche et al.,  
502 2021). For metaproteomic analysis, this process is even more essential, as variations in results  
503 have been detected depending on the used workflow even in modern metaproteomic studies (Van  
504 Den Bossche et al., 2021).

505

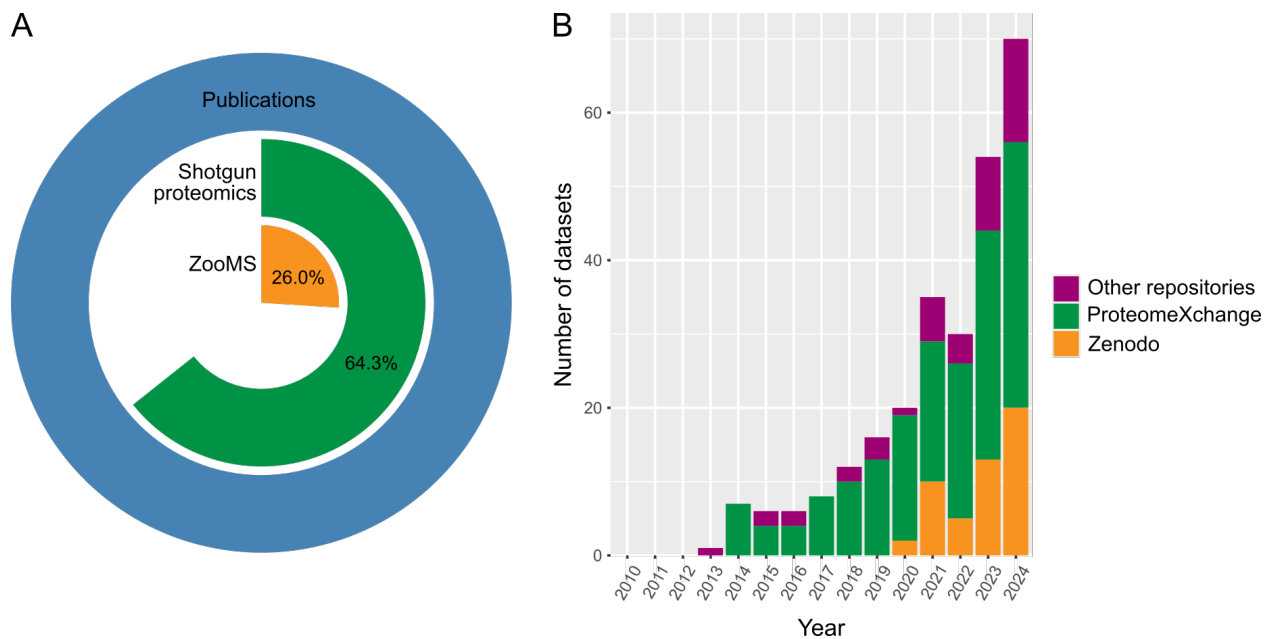
506 Archaeological sciences are increasingly utilising code-based analysis methods tailored  
507 to specific research requirements, such as workflows written in R (R Core Team, 2023) or Python.  
508 This allows for optimisation of analyses for each specific project's needs, automation of processes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

519  
520 Although palaeoproteomic analyses can provide novel insights about our past, the  
521 proteomic results most often cannot stand on their own. For example, without dating and  
522 archaeological context, the taxonomic identification of a specimen cannot be meaningfully  
523 interpreted. Similarly, identification of pathogen or immune-related proteins should have a basis  
524 in clinical studies to ensure that conclusions are based on clinically reliable evidence. It is  
525 therefore essential to collaborate with researchers from other fields, from the very beginning of a  
526 project, to ensure that all available information is combined into a holistic interpretation of the  
527 proteomic results. Archaeogenetic and palaeoproteomic analyses should not be seen as  
528 competing, but rather as complementary. For example, genetic sequences contain more  
529 phylogenetic information than protein sequences, but DNA degrades at a higher rate than some  
530 proteins, and the two fields therefore have different points of strength. They may also have  
531 differing strengths in the type of information they can provide. For example, in studies of  
532 archaeological dental calculus, genetic studies have been more promising for studying the oral  
533 microbiome (Fellows Yates et al., 2021; Ottoni et al., 2021; Warinner, Hendy, et al., 2014),  
534 whereas proteomics has provided more information about past diets (Scott et al., 2021; Warinner,  
535 Hendy, et al., 2014; Wilkin et al., 2021). Additionally, other emerging fields such as  
536 palaeometabolomics (Barberis et al., 2022; Brownstein et al., 2020; Huber et al., 2023; Velsko et  
537 al., 2017) can be integrated for a comprehensive biomolecular toolkit for analysis of  
538 archaeological materials.

## 539 Data Sharing and Metadata Availability

540  
541 The consensus to make data openly available has strengthened and driven development of  
542 bioinformatic methods in many fields (e.g., Anagnostou et al., 2015; Celi et al., 2019; Lapatas et  
543 al., 2015). This development seems to still be in progress in the field of palaeoproteomics,  
544 whereby data is being produced and analysed but not always being made available alongside  
545 those publications (Fig. 2). This issue is most prominent for ZooMS studies, where raw data is  
546 only available for 26% of all studies that produced data (Fig. 2A). The picture is more positive for  
547 shotgun palaeoproteomics, with data being available for 64.3% of studies that produced data.

548 Despite these low percentages, we do see an increase in publications with available data in recent  
 549 years (Fig. 2B), most likely due to the concurrent increase in palaeoproteomics studies in recent  
 550 years, and perhaps a response to an earlier call for the public sharing of raw and processed data  
 551 by Hendy *et al.* (Hendy, Welker, et al., 2018). Although the problem of low reporting of raw data  
 552 for ZooMS manuscripts has been previously discussed (Richter et al., 2022), there is currently no  
 553 agreed-upon standard for what data to publish, or how, alongside ZooMS manuscripts. We aim  
 554 for the PAASTA community to facilitate the development of data publishing standards for both  
 555 ZooMS and shotgun palaeoproteomics datasets, and we also hope to encourage researchers to  
 556 revisit their prior work and, where possible, make that data available. It is worth noting that it is  
 557 never too late to make data from prior publications available. We do, however, acknowledge that  
 558 there might be factors that complicate open data sharing, for example, the importance of  
 559 indigenous communities holding stewardship over their own data (Carroll et al., 2020).  
 560



561  
 562 *Figure 2. Data availability in palaeoproteomics. A) Percentage of ZooMS and shotgun palaeoproteomics*  
 563 *studies with available datasets. Preprints and papers that did not produce any data were excluded from this*  
 564 *calculation. B) Number of palaeoproteomics datasets in online repositories (since 2010). ‘Other’*  
 565 *repositories include Dryad, Figshare, Github, Mendeley Data, and the OSF Framework. The dataset was*  
 566 *compiled by manually going through the list of publications compiled for Figure 1 and checking raw data*  
 567 *availability. This figure was made in R v.4.4.2. with R Studio v.2024.12.0+467 (R Core Team, 2023) using*  
 568 *ggplot2 (Wickham, 2016).*

569  
 570 Equally important as open data sharing is the appropriate description of metadata. If the  
 571 nature of the sample or the relationship between the analysed data and the samples is not clear,  
 572 the reusability of data will decrease or disappear. The MAGE-TAB-Proteomics format (Dai et al.,  
 573 2021) has been proposed by ProteomeXchange as a metadata format (Deutsch et al., 2023).  
 574 MAGE-TAB-Proteomics has two components: the Investigation Description Format (IDF) and the  
 575 Sample and Data Relationship Format (SDRF-Proteomics, updated in (Claeys et al., 2023). The  
 576 IDF, which is a general description of the research, is automatically generated from the mandatory

577 information for registration in the ProteomeXchange repository, but the SDRF-Proteomics, which  
578 describes the details of the samples and the experiments, needs to be created and registered by  
579 the researcher. Since SDRF-Proteomics was not designed for archaeological materials, some  
580 information, such as the age or region of origin, specific to archaeological or paleontological  
581 materials, lacks a corresponding data input column. It may thus be necessary to create a revised  
582 format for archeological/paleontological material-specific metadata, as has been necessary in the  
583 field of archaeogenetics (Bergström et al., 2024).

584  
585 To aid in alleviating these concerns going forward and in an effort to establish best  
586 practices in the field, we propose the following as minimum reporting standards for  
587 palaeoproteomics, additionally adhering to the FAIR and CARE principles:

- 588 ● When LC-MS/MS data is generated, the raw files should be made available in a public  
589 data repository. This should include raw data files from all analysed specimens, as well  
590 as extraction blanks. Our recommendation is to submit these files to ProteomeXchange,  
591 to which data can be deposited via several platforms including PRIDE and MassIVE.  
592 Submissions should ideally include the required files for a complete submission as  
593 detailed in their guidelines.
- 594 ● Reconstructed protein sequences should be shared as FASTA files in supplementary data  
595 with the manuscript, or uploaded alongside raw data files to a repository such as  
596 ProteomeXchange. Note that GenBank and other protein data repositories do not accept  
597 submissions of reconstructed partial protein sequences derived from palaeoproteomics  
598 research, as there is no evidence for the length of the unknown sequence regions. Protein  
599 sequences should not be shared in .pdf format as this is a less accessible method of  
600 sharing.
- 601 ● For MALDI-ToF or MALDI-FT-ICR data (for ZooMS), the raw data should be made  
602 available in a public data repository. Ideally, files with the processed data should also be  
603 made available for the sake of reproducibility. Our recommended repository for this data  
604 is Zenodo in .mzXML or .txt file format, or ProteomeXchange, as .mzXML files can be  
605 uploaded.
- 606 ● Where code has been used to analyse data, the code should be provided so that analyses  
607 can be replicated and results/interpretations are transparent. We recommend sharing  
608 code through a version-tracked repository system such as GitHub or Zenodo, and we  
609 recommend that the code be archived using a persistent identifier, such as a DOI.
- 610 ● Appropriate and relevant metadata of samples should be included in the form of a  
611 metadata file, such as SDRF-Proteomics, highlighting information such as:
  - 612 ○ Ancient and modern samples: links between samples and raw file names/spectra,  
613 sample IDs, batch and plate numbers, method of protein extraction, details of  
614 analytical machines used (including name of instrument and vendor) and precise  
615 parameters employed.
  - 616 ○ Ancient samples: context of the samples, dating method of samples and  
617 corresponding time period, notes of any treatments post-excavation.
  - 618 ○ Modern samples: source of reference materials, details of how species were  
619 identified, and if the sample was macerated, include notes on how this was done  
620 and how the sample has been stored and treated.

- 621 ● All of the shared files should be in an accessible format, from which data can be readily  
622 extracted (i.e., not PDFs, images, etc.).
- 623 ● Ideally, research should be published in an Open Access format, but at a minimum,  
624 authors should place accepted versions in institutional repositories to meet public access  
625 requirements. However, we are aware of the, at times, prohibitively expensive costs of  
626 publishing Open Access in numerous journals. There is an increasing number of Open  
627 Access journals, as well as alternative publishing platforms, which can circumnavigate this  
628 issue. Ensuring that the corresponding data is made publicly available, even if the  
629 manuscript is not, is recommended and in spirit with our recommended best practices.

## 630 Future Perspectives

631  
632 The field of palaeoproteomics has advanced significantly over the past years, thanks to a range  
633 of emerging analytical and computational methodologies. Such innovations have enhanced the  
634 sensitivity and accuracy of ancient protein analysis, thereby providing deeper insights into the  
635 past. With these innovations leading to an ever-growing number of researchers and research  
636 applications, we are advocating to advance palaeoproteomics through keeping open science,  
637 collaboration and communication at the center of our research. These practices will enable our  
638 growing community of researchers to validate and build on each other's work, collaborate more  
639 effectively, and ultimately accelerate scientific progress in palaeoproteomics. We additionally  
640 hope that this paper can serve as a starting point for our collaborators outside of palaeoproteomics  
641 (e.g., curators and archaeologists) to understand what standards and approaches our community  
642 holds as best practices.

643  
644 Future developments in both mass spectrometry and data processing are likely to further  
645 transform the field in a multitude of ways. Mass spectrometry instrumentation used in  
646 palaeoproteomics is crucial for advancing the field, although often accompanied by considerable  
647 costs and the need for proprietary software. Recently, new technologies for single-molecule  
648 proteomics that do not necessarily rely on mass spectrometry, such as nanopores or molecular  
649 markers, are being developed (Alfaro et al., 2021; Floyd & Marcotte, 2022), and these new  
650 technologies have the potential to be applied to archaeological and paleontological samples  
651 (Alfaro et al., 2021). Frequent incremental upgrades in instrumentation and software or drastic  
652 technological innovation are often necessary to stay at the forefront of research, posing a potential  
653 barrier to widespread method adoption. However, continued investment in state-of-the-art  
654 instrumentation is essential for advancing palaeoproteomics research. Artificial intelligence (AI)  
655 systems have revolutionised protein tertiary structure prediction (e.g., Jumper et al., 2021;  
656 Tunyasuvunakool et al., 2021), which may impact the future of palaeoproteomics. In  
657 palaeoproteomic research, where recovered protein sequences are highly degraded, AI systems  
658 may aid researchers in inferring protein structure, thereby improving understanding in  
659 evolutionary relationships and functionality (Demarchi et al., 2022), as well as potentially enhance  
660 the annotation and characterisation of ancient proteins. The use of AI in palaeoproteomics is likely  
661 to significantly advance the identification and analysis of proteins and proteomes (Chiang &  
662 Collins, 2024), but will need to be extensively validated. These types of major advances can be

663 efficiently evaluated and adapted into the field where appropriate, through collaboration between  
664 researchers from different subfields of palaeoproteomics.

665  
666 Importantly, many of the challenges, opportunities and problems outlined above can be  
667 overcome by communication between researchers and research groups. This is true both within  
668 the field of palaeoproteomics, as well as between research fields. Working together on projects  
669 (such as the creation of an open access, community-curated database of ZooMS peptide  
670 markers), software, and applications ensures that advances—and failures—are properly  
671 communicated. Such collaborative projects can also be used by the community to steer the  
672 direction of method development. Experiments that are valuable to the community, but are not  
673 easily funded by individual grants, such as reproducing previous results, validation of new  
674 methods or benchmarking tests, may become feasible if shared by many researchers from  
675 different institutions. Especially in the case of software validation, for which it can be prohibitively  
676 time consuming for the developer to test on a wide range of sample sets, a community-based  
677 approach may provide the large scale of diverse datasets needed and share the time-load  
678 required to validate new tools for palaeoproteomic applications.

679  
680 Currently, a lot of palaeoproteomics expertise on the use of various extraction protocols,  
681 mass spectrometry settings and software is limited to a small number of individuals and  
682 institutions, and detailed (and practical) knowledge of palaeoproteomics is rarely taught at  
683 universities. As the palaeoproteomics community grows and the number of institutes with  
684 palaeoproteomics researchers increases, the importance of sharing knowledge and experiences  
685 increases as well. Not only will facilitating the sharing of technical know-how make  
686 palaeoproteomics more accessible to new researchers, it will also ensure that community  
687 conventions and high research standards are maintained, and provide researchers with an  
688 overview of the range of research tools available. Building a community such as PAASTA  
689 provides researchers with a platform for such conversations, allowing networking, collaboration,  
690 method development and validation, and knowledge sharing.

## 691 Conclusions

692  
693 The PAASTA community aims to create an open and supportive environment for  
694 palaeoproteomics researchers to share knowledge and expertise, and to foster collaborations and  
695 open science. We have written this manuscript as a means to establish and renew standards for  
696 best practice in palaeoproteomics research, emphasising data sharing and reporting, and  
697 promoting open science among the rapidly growing community of researchers within this field.  
698 Our key recommendations and proposed guidelines are as follows:

- 699 ○ Open and collaborative discussions with various stakeholders in cultural and  
700 natural heritage regarding sample selection, documentation, planned proteomic  
701 analyses and interpretation.
- 702 ○ Detailed and reproducible protocols, as well as increased sharing of negative  
703 results.

- 704           ○ Development of analysis methods and pipelines in a community-supported  
705           manner, and detailed sharing of analysis methods and parameters.  
706           ○ Raw data, protein databases, and relevant code to be made available alongside  
707           publications and in an accessible format and provide appropriate accompanying  
708           detailed metadata.

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

715

## 716 Acknowledgements

717

718 We would like to thank the entire PAASTA community for discussions and collaborations which  
719 have led to this manuscript. This research has been made possible through funding from the  
720 European Research Council (ERC) under the European Union's Horizon 2020 research and  
721 innovation programme, grant agreement no. 787282 (Beasts2Craft, supporting M.C. and B.N.),  
722 no. 948365 (PROSPER, awarded to F.W.), and no. 865222 (EQuaTe, supporting M.R.D.), as well  
723 as under the Marie Skłodowska-Curie grant agreement no. 956351 (ChemArch, supporting J.D.),  
724 and the European Union's Horizon Europe research and innovation programme under the Marie  
725 Skłodowska-Curie grant agreement no. 101106627 (PROMISE, awarded to Z.F.), and ERC grant  
726 agreement no. 101041245 (MATRIX, supporting C.P.) and no. 101039416 (Fashioning Sudan,  
727 supporting L.C.V.C.), and the Volkswagen Foundation (C.W.), and by the Deutsche  
728 Forschungsgemeinschaft (460129525, NFDI4 Microbiota EnterArchaeo, C.W.), and the Werner  
729 Siemens Foundation (Paleobiotechnology, C.W.), and the U.S. National Science Foundation  
730 (C.W.), and the Harvard Radcliffe Institute (C.W.). T.T. is supported by the Japan Science and  
731 Technology Agency (FOREST, no. JPMJFR233D). R.M.W. is supported by the French National  
732 Research Agency (ANR) Young Researcher (JCJC) Grant awarded to Sophia V. Hansson (ANR-  
733 21-CE34-0001, ATCAF-project). F.W. is supported by VILLUM FONDEN (no. 40747). Views and  
734 opinions expressed are however those of the author(s) only and do not necessarily reflect those  
735 of the European Union, the European Research Council Executive Agency, or any of the other  
736 funders. Neither the European Union nor the granting agency can be held responsible for them.  
737 We additionally thank the reviewers for their valuable suggestions which improved the manuscript.

738

## 739 Code and data availability

740 R code and supporting data have been archived at [10.5281/zenodo.14899604](https://zenodo.org/record/14899604).

741

## 742 References

- 743 Abelson, P. H. (1954). Amino acids in fossils. *Science*, 119(3096), 576.
- 744 Alfaro, J. A., Bohländer, P., Dai, M., Filius, M., Howard, C. J., van Kooten, X. F., Ohayon, S.,  
745 Pomorski, A., Schmid, S., Aksimentiev, A., Anslyn, E. V., Bedran, G., Cao, C., Chinappi,  
746 M., Coyaud, E., Dekker, C., Dittmar, G., Drachman, N., Eelkema, R., ... Joo, C. (2021). The  
747 emerging landscape of single-molecule protein sequencing technologies. *Nature Methods*,  
748 18(6), 604–617.
- 749 Alves, R. D. A. M., Demmers, J. A. A., Bezstarosti, K., van der Eerden, B. C. J., Verhaar, J. A.  
750 N., Eijken, M., & van Leeuwen, J. P. T. M. (2011). Unraveling the human bone  
751 microenvironment beyond the classical extracellular matrix proteins: a human bone protein  
752 library. *Journal of Proteome Research*, 10(10), 4725–4733.
- 753 Anagnostou, P., Capocasa, M., Milia, N., Sanna, E., Battaggia, C., Luzi, D., & Destro Bisol, G.  
754 (2015). When data sharing gets close to 100%: what human paleogenetics can teach the  
755 open science movement. *PLoS One*, 10(3), e0121409.
- 756 Ásmundsdóttir, R. D., Hansen, J., Fagernäs, Z., Troché, G., Olsen, J. V., Saña Seguí, M., &  
757 Welker, F. (2024). Early Holocene preservation differences between cortical and trabecular  
758 bone proteomes. *Journal of Archaeological Science: Reports*, 57, 104643.
- 759 Azémard, C., Dufour, E., Zazzo, A., Wheeler, J. C., Goepfert, N., Marie, A., & Zirah, S. (2021).  
760 Untangling the fibre ball: Proteomic characterization of South American camelid hair fibres  
761 by untargeted multivariate analysis and molecular networking. *Journal of Proteomics*, 231,  
762 104040.
- 763 Bache, N., Geyer, P. E., Bekker-Jensen, D. B., Hoerning, O., Falkenby, L., Treit, P. V., Doll, S.,  
764 Paron, I., Müller, J. B., Meier, F., Olsen, J. V., Vorm, O., & Mann, M. (2018). A novel LC  
765 system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics.  
766 *Molecular & Cellular Proteomics*, 17(11), 2284–2296.
- 767 Baker, A., Harvey, V. L., & Buckley, M. (2023). Machine Learning for collagen peptide biomarker  
768 determination in the taxonomic identification of archaeological fish remains. *Journal of*  
769 *Archaeological Science: Reports*, 49, 104001.
- 770 Barberis, E., Manfredi, M., Ferraris, E., Bianucci, R., & Marengo, E. (2022). Non-invasive Paleo-  
771 metabolomics and Paleo-proteomics analyses reveal the complex funerary treatment of the  
772 early 18th Dynasty dignitary NEBIRI (QV30). *Molecules*, 27(21), 7208.
- 773 Barker, A., Venables, B., Stevens, S. M., Seeley, K. W., Wang, P., & Wolverton, S. (2012). An  
774 optimized approach for protein residue extraction and identification from ceramics after  
775 cooking. *Journal of Archaeological Method and Theory*, 19(3), 407–439.
- 776 Barnard, H., Ambrose, S. H., Beehr, D. E., Forster, Lanehart, R. E., Malainey, M. E., Parr, R. E.,  
777 Rider, M., Solazzo, C., & Yohe, R. M. (2007). Mixed results of seven methods for organic  
778 residue analysis applied to one vessel with the residue of a known foodstuff. *Journal of*  
779 *Archaeological Science*, 34(1), 28–37.
- 780 Bell, P. A., Solis, N., Kizhakkedathu, J. N., Matthew, I., & Overall, C. M. (2019). Proteomic and  
781 N-terminomic TAILS analyses of human alveolar bone proteins: Improved protein extraction  
782 methodology and LysargiNase digestion strategies increase proteome coverage and  
783 missing protein identification. *Journal of Proteome Research*, 18(12), 4167–4179.
- 784 Bergström, A., Fellows Yates, J. A., & Warinner, C. (2024). Ancient DNA data hold insights into  
785 past organisms and ecosystems - handle them with more care. *Nature*, 636(8042), 296–  
786 298.
- 787 Bern, M., Kil, Y. J., & Becker, C. (2012). Byonic: advanced peptide and protein identification  
788 software. *Current Protocols in Bioinformatics* 13(1), 13.20.1–13.20.14.
- 789 Brandt, L. Ø., Lillemark, M. R., Rytter, M., Collins, M. J., & Tøttrup, A. P. (2022). Next  
790 Generation Lab puts Denmark's past in the hands of its future. *Antiquity*, 96(385), 221–228.
- 791 Brandt, L. Ø., Schmidt, A. L., Mannering, U., Sarret, M., Kelstrup, C. D., Olsen, J. V., &

792 Cappellini, E. (2014). Species identification of archaeological skin objects from Danish  
793 bogs: comparison between mass spectrometry-based peptide sequencing and microscopy-  
794 based methods. *PLoS One*, 9(9), e106875.

795 Brazil, R. (2024). Illuminating “the ugly side of science”: fresh incentives for reporting negative  
796 results. *Nature Career Feature*.

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

798 Brown, S., Douka, K., Collins, M. J., & Richter, K. K. (2021). On the standardization of ZooMS  
799 nomenclature. *Journal of Proteomics*, 235, 104041.

800 Brown, S., Hebestreit, S., Wang, N., Boivin, N., Douka, K., & Kozow Richter, K. (2020).  
801 Zooarchaeology by Mass Spectrometry (ZooMS) for bone material - AmbIC protocol.  
802 *Protocols.io*.

803 Brown, S., Kozlikin, M., Shunkov, M., Derevianko, A., Higham, T., Douka, K., & Richter, K. K.  
804 (2021). Examining collagen preservation through glutamine deamidation at Denisova Cave.  
805 *Journal of Archaeological Science*, 133, 105454.

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

809 Buckley, M. (2015). Ancient collagen reveals evolutionary history of the endemic South  
810 American “ungulates.” *Proceedings of the Royal Society B: Biological Sciences*, 282(1806),  
811 20142671.

812 Buckley, M., Collins, M., Thomas-Oates, J., & Wilson, J. C. (2009). Species identification by  
813 analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight  
814 mass spectrometry. *Rapid Communications in Mass Spectrometry*, 23(23), 3843–3854.

815 Buckley, M., Fraser, S., Herman, J., Melton, N. D., Mulville, J., & Pálsdóttir, A. H. (2014).  
816 Species identification of archaeological marine mammals using collagen fingerprinting.  
817 *Journal of Archaeological Science*, 41, 631–641.

818 Callaway, E. (2017). Stop hoarding ancient bones, plead archaeologists. *Nature News*.

819 Cappellini, E., Jensen, L. J., Szklarczyk, D., Ginolhac, A., da Fonseca, R. A. R., Stafford, T. W.,  
820 Holen, S. R., Collins, M. J., Orlando, L., Willerslev, E., Gilbert, M. T. P., & Olsen, J. V.  
821 (2012). Proteomic analysis of a pleistocene mammoth femur reveals more than one  
822 hundred ancient bone proteins. *Journal of Proteome Research*, 11(2), 917–926.

823 Cappellini, E., Welker, F., Pandolfi, L., Ramos-Madrugal, J., Samodova, D., Rütther, P. L.,  
824 Fotakis, A. K., Lyon, D., Moreno-Mayar, J. V., Bukhsianidze, M., Rakownikow Jersie-  
825 Christensen, R., Mackie, M., Ginolhac, A., Ferring, R., Tappen, M., Palkopoulou, E.,  
826 Dickinson, M. R., Stafford, T. W., Jr, Chan, Y. L., ... Willerslev, E. (2019). Early Pleistocene  
827 enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature*, 574(7776),  
828 103–107.

829 Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S.,  
830 Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D.,  
831 Anderson, J., & Hudson, M. (2020). The CARE principles for indigenous data governance.  
832 *Data Science Journal*, 19(1).

833 Celi, L. A., Citi, L., Ghassemi, M., & Pollard, T. J. (2019). The PLOS ONE collection on machine  
834 learning in health and biomedicine: Towards open code and open data. *PLoS One*, 14(1),  
835 e0210232.

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

838 Chen, F., Welker, F., Shen, C.-C., Bailey, S. E., Bergmann, I., Davis, S., Xia, H., Wang, H.,  
839 Fischer, R., Freidline, S. E., Yu, T.-L., Skinner, M. M., Stelzer, S., Dong, G., Fu, Q., Dong,  
840 G., Wang, J., Zhang, D., & Hublin, J.-J. (2019). A late Middle Pleistocene Denisovan  
841 mandible from the Tibetan Plateau. *Nature*, 569(7756), 409–412.

842 Chi, H., Liu, C., Yang, H., Zeng, W.-F., Wu, L., Zhou, W.-J., Wang, R.-M., Niu, X.-N., Ding, Y.-

843 H., Zhang, Y., Wang, Z.-W., Chen, Z.-L., Sun, R.-X., Liu, T., Tan, G.-M., Dong, M.-Q., Xu,  
844 P., Zhang, P.-H., & He, S.-M. (2018). Comprehensive identification of peptides in tandem  
845 mass spectra using an efficient open search engine. *Nature Biotechnology*, 36(11), 1059–  
846 1061.

847 Chiang, Y., & Collins, M. J. (2024). Orthrus: an AI-powered, cloud-ready, and open-source  
848 hybrid approach for metaproteomics. *bioRxiv* 2024.11.15.623814.

849 Claeys, T., Van Den Bossche, T., Perez-Riverol, Y., Gevaert, K., Vizcaíno, J. A., & Martens, L.  
850 (2023). lesSDRF is more: maximizing the value of proteomics data through streamlined  
851 metadata annotation. *Nature Communications*, 14(1), 6743.

852 Cleland, T. P., Schroeter, E. R., Feranec, R. S., & Vashishth, D. (2016). Peptide sequences  
853 from the first *Castoroides ohioensis* skull and the utility of old museum collections for  
854 palaeoproteomics. *Proceedings of the Royal Society B: Biological Sciences*, 283(1832).

855 Cleland, T. P., Schroeter, E. R., & Schweitzer, M. H. (2015). Biologically and diagenetically  
856 derived peptide modifications in moa collagens. *Proceedings of the Royal Society B:  
857 Biological Sciences*, 282(1808), 20150015.

858 Codlin, M. C., Douka, K., & Richter, K. K. (2022). An application of zooms to identify  
859 archaeological avian fauna from Teotihuacan, Mexico. *Journal of Archaeological Science*,  
860 148, 105692.

861 Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized  
862 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature  
863 Biotechnology*, 26(12), 1367–1372.

864 Dai, C., Füllgrabe, A., Pfeuffer, J., Solovyeva, E. M., Deng, J., Moreno, P., Kamatchinathan, S.,  
865 Kundu, D. J., George, N., Fexova, S., Grüning, B., Föll, M. C., Griss, J., Vaudel, M., Audain,  
866 E., Locard-Paulet, M., Turewicz, M., Eisenacher, M., Uszkoreit, J., ... Perez-Riverol, Y.  
867 (2021). A proteomics sample metadata representation for multiomics integration and big  
868 data analysis. *Nature Communications*, 12(1), 5854.

869 Dekker, J. A. A., Hagan, R., Collins, M. J., & Hendy, J. (2025). An unclean slate, discrepancies  
870 between food input and recovered protein signal from experimental foodcrusts. *bioRxiv*  
871 2025.02.03.636240.

872 Demarchi, B., Stiller, J., Grealy, A., Mackie, M., Deng, Y., Gilbert, T., Clarke, J., Legendre, L. J.,  
873 Boano, R., Sicheritz-Pontén, T., Magee, J., Zhang, G., Bunce, M., Collins, M. J., & Miller,  
874 G. (2022). Ancient proteins resolve controversy over the identity of *Genyornis* eggshell.  
875 *Proceedings of the National Academy of Sciences of the United States of America*,  
876 119(43), e2109326119.

877 Demeter, F., Zanolli, C., Westaway, K. E., Joannes-Boyau, R., Düringer, P., Morley, M. W.,  
878 Welker, F., Rütther, P. L., Skinner, M. M., McColl, H., Gaunitz, C., Vinner, L., Dunn, T. E.,  
879 Olsen, J. V., Sikora, M., Ponche, J.-L., Suzzoni, E., Frangeul, S., Boesch, Q., ...  
880 Shackelford, L. (2022). A Middle Pleistocene Denisovan molar from the Annamite Chain of  
881 northern Laos. *Nature Communications*, 13(1), 2557.

882 Deutsch, E. W., Bandeira, N., Perez-Riverol, Y., Sharma, V., Carver, J. J., Mendoza, L., Kundu,  
883 D. J., Wang, S., Bandla, C., Kamatchinathan, S., Hewapathirana, S., Pullman, B. S., Wertz,  
884 J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., MacLean, B., MacCoss, M. J., ...  
885 Vizcaíno, J. A. (2023). The ProteomeXchange consortium at 10 years: 2023 update.  
886 *Nucleic Acids Research*, 51(D1), D1539–D1548.

887 Di Gianvincenzo, F., Andersen, C. K., Filtenborg, T., Mackie, M., Ernst, M., Ramos Madrigal, J.,  
888 Olsen, J. V., Wadum, J., & Cappellini, E. (2023). Proteomic identification of beer brewing  
889 products in the ground layer of Danish Golden Age paintings. *Science Advances*, 9(21),  
890 eade7686.

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

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

898 Ebsen, J. A., Haase, K., Larsen, R., Sommer, D. V. P., & Brandt, L. Ø. (2019). Identifying  
899 archaeological leather – discussing the potential of grain pattern analysis and  
900 zooarchaeology by mass spectrometry (ZooMS) through a case study involving medieval  
901 shoe parts from Denmark. *Journal of Cultural Heritage*, 39, 21–31.

902 Eckhardt, A., Jágr, M., Pataridis, S., & Mikšík, I. (2014). Proteomic analysis of human tooth pulp:  
903 proteomics of human tooth. *Journal of Endodontia*, 40(12), 1961–1966.

904 Engels, I., Burnett, A., Robert, P., Pironneau, C., Abrams, G., Bouwmeester, R., Van der  
905 Plaetsen, P., Di Modica, K., Otte, M., Straus, L. G., Fischer, V., Bray, F., Mesuere, B., De  
906 Groote, I., Deforce, D., Daled, S., & Dhaenens, M. (2025). Classification of collagens via  
907 peptide ambiguity, in a paleoproteomic LC-MS/MS-based taxonomic pipeline. *Journal of*  
908 *Proteome Research*, 24(4), 1907–1925

909 Evans, M., Hagan, R., Boyd, O., Bondetti, M., Craig, O., Collins, M. J., & Hendy, J. (2024). The  
910 impact of cooking and burial on proteins: a characterisation of experimental foodcrusts and  
911 ceramics. *Royal Society Open Science*, 11(240610).

912 Evans, Z., Paskulin, L., Rahemtulla, F., & Speller, C. F. (2023). A comparison of minimally-  
913 invasive sampling techniques for ZooMS analysis of bone artifacts. *Journal of*  
914 *Archaeological Science: Reports*, 47(103738), 103738.

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

918 Fagernäs, Z., Troché, G., Olsen, J. V., & Welker, F. (2024). Digging deeper into ancient skeletal  
919 proteomes through consecutive digestion with multiple proteases. *Journal of Proteomics*,  
920 298, 105143.

921 Fagernäs, Z., Villa-Islas, V., Troché, G., Buylaert, J.-P., Khujageldiev, T., Kurbanov, R., Olsen,  
922 J. V., Pedersen, M. W., & Welker, F. (2025). Cleaning the dead: Optimized  
923 decontamination enhances palaeoproteomic analyses of a Pleistocene hominin tooth from  
924 Khudji, Tajikistan. *Journal of Archaeological Science*, 106228, 106228.

925 Fellows Yates, J. A., Velsko, I. M., Aron, F., Posth, C., Hofman, C. A., Austin, R. M., Parker, C.  
926 E., Mann, A. E., Nägele, K., Arthur, K. W., Arthur, J. W., Bauer, C. C., Crevecoeur, I.,  
927 Cupillard, C., Curtis, M. C., Dalén, L., Díaz-Zorita Bonilla, M., Díez Fernández-Lomana, J.  
928 C., Drucker, D. G., ... Warinner, C. (2021). The evolution and changing ecology of the  
929 African hominid oral microbiome. *Proceedings of the National Academy of Sciences of the*  
930 *United States of America*, 118(20).

931 Fiddyment, S., Holsinger, B., Ruzzier, C., Devine, A., Binois, A., Albarella, U., Fischer, R.,  
932 Nichols, E., Curtis, A., Cheese, E., Teasdale, M. D., Checkley-Scott, C., Milner, S. J., Rudy,  
933 K. M., Johnson, E. J., Vnouček, J., Garrison, M., McGrory, S., Bradley, D. G., & Collins, M.  
934 J. (2015). Animal origin of 13th-century uterine vellum revealed using noninvasive peptide  
935 fingerprinting. *Proceedings of the National Academy of Sciences of the United States of*  
936 *America*, 112(49), 15066–15071.

937 Fiddyment, S., Teasdale, M. D., Vnouček, J., Lévêque, É., Binois, A., & Collins, M. J. (2019). So  
938 you want to do biocodicology? A field guide to the biological analysis of parchment.  
939 *Heritage Science*, 7(1), 1–10.

940 Fleskes, R. E., Bader, A. C., Tsoie, K. S., Wagner, J. K., Claw, K. G., & Garrison, N. A. (2022).  
941 Ethical guidance in human paleogenomics: New and ongoing perspectives. *Annual Review*  
942 *of Genomics and Human Genetics*, 23(1), 627–652.

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

945 Fonseca, B., Freeman, C. L., & Collins, M. J. (2024). The survival of  $\beta$ -lactoglobulin peptides in  
946 the archaeological record: Vulnerability vs. Sequence variation. *bioRxiv*  
947 2024.09.13.612646.

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

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

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

957 Haghighi, Z., Mackie, M., Apalnes Ørnhøi, A., Ramsøe, A., Olstad, T. M., Armitage, S. J.,  
958 Henshilwood, C. S., & Cappellini, E. (2024). Palaeoproteomic identification of the original  
959 binder and modern contaminants in distemper paints from Uvdal stave church, Norway.  
960 *Scientific Reports*, 14(1), 12858.

961 Hansen, J., Dekker, J., Troché, G., Fagernäs, Z., Olsen, J. V., Saña Seguí, M., & Welker, F.  
962 (2024). A comparative study of commercially available, minimally invasive, sampling  
963 methods on Early Neolithic humeri analysed via palaeoproteomics. *Journal of*  
964 *Archaeological Science*, 167(106002), 106002.

965 Harvey, V. L., Daugnora, L., & Buckley, M. (2018). Species identification of ancient Lithuanian  
966 fish remains using collagen fingerprinting. *Journal of Archaeological Science*, 98, 102–111.

967 Harvey, V. L., Keating, J. N., & Buckley, M. (2021). Phylogenetic analyses of ray-finned fishes  
968 (Actinopterygii) using collagen type I protein sequences. *Royal Society Open Science*, 8(8),  
969 201955.

970 Haug, C., Reumer, J. W. F., Haug, J. T., Arillo, A., Audo, D., Azar, D., Baranov, V., Beutel, R.,  
971 Charbonnier, S., Feldmann, R., Foth, C., Fraaije, R. H. B., Frenzel, P., Gašparič, R.,  
972 Greenwalt, D. E., Harms, D., Hyžný, M., Jagt, J. W. M., Jagt-Yazykova, E. A., ... Reich, M.  
973 (2020). Comment on the letter of the Society of Vertebrate Paleontology (SVP) dated April  
974 21, 2020 regarding “Fossils from conflict zones and reproducibility of fossil-based scientific  
975 data”: the importance of private collections. *Palaontologische Zeitschrift*, 94(3), 413–429.

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

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

982 Hendy, J., Welker, F., Demarchi, B., Speller, C., Warinner, C., & Collins, M. J. (2018). A guide to  
983 ancient protein studies. *Nature Ecology & Evolution*, 2(5), 791–799.

984 Hickinbotham, S., Fiddyment, S., Stinson, T. L., & Collins, M. J. (2020). How to get your goat:  
985 automated identification of species from MALDI-ToF spectra. *Bioinformatics*, 36(12), 3719–  
986 3725.

987 Hollemeyer, K., Altmeyer, W., Heinzle, E., & Pitra, C. (2008). Species identification of Oetzi's  
988 clothing with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry  
989 based on peptide pattern similarities of hair digests. *Rapid Communications in Mass*  
990 *Spectrometry*, 22(18), 2751–2767.

991 Horsburgh, A. K. (2024). Concluding Commentary. In D. Strand, A. Källén, & C. Mulcare (Eds.),  
992 *Critical Perspectives on ancient DNA* (pp. 167–175). The MIT Press.

993 Huber, B., Hammann, S., Loeben, C. E., Jha, D. K., Vassão, D. G., Larsen, T., Spengler, R. N.,  
994 Fuller, D. Q., Roberts, P., Deviese, T., & Boivin, N. (2023). Biomolecular characterization of  
995 3500-year-old ancient Egyptian mummification balms from the Valley of the Kings.

996 *Scientific Reports*, 13(1), 12477.

997 Jágr, M., Ergang, P., Pataridis, S., Kolrosová, M., Bartoš, M., & Mikšík, I. (2019). Proteomic  
998 analysis of dentin-enamel junction and adjacent protein-containing enamel matrix layer of  
999 healthy human molar teeth. *European Journal of Oral Sciences*, 127(2), 112–121.

1000 Janzen, A., Richter, K. K., Mwebi, O., Brown, S., Onduso, V., Gatwiri, F., Ndiema, E., Katongo,  
1001 M., Goldstein, S. T., Douka, K., & Boivin, N. (2021). Distinguishing African bovids using  
1002 Zooarchaeology by Mass Spectrometry (ZooMS): New peptide markers and insights into  
1003 Iron Age economies in Zambia. *PloS One*, 16(5), e0251061.

1004 Jersie-Christensen, R. R., Lanigan, L. T., Lyon, D., Mackie, M., Belstrøm, D., Kelstrup, C. D.,  
1005 Fotakis, A. K., Willerslev, E., Lynnerup, N., Jensen, L. J., Cappellini, E., & Olsen, J. V.  
1006 (2018). Quantitative metaproteomics of medieval dental calculus reveals individual oral  
1007 health status. *Nature Communications*, 9(1), 4744.

1008 Jones, E. D. (2022). *Ancient DNA: The Making of a Celebrity Science*. Yale University Press.

1009 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool,  
1010 K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A.  
1011 J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021).  
1012 Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.

1013 Källén, A. (2025). *The Trouble with Ancient DNA*. University of Chicago Press.

1014 Källén, A., Mulcare, C., Nyblom, A., & Strand, D. (2024). Petrous fever: The gap between ideal  
1015 and actual practice in ancient DNA research. *Current Anthropology*, 65(6), 990-1009..

1016 Kirby, D. P., Buckley, M., Promise, E., Trauger, S. A., & Holdcraft, T. R. (2013). Identification of  
1017 collagen-based materials in cultural heritage. *The Analyst*, 138(17), 4849–4858.

1018 Kirby, D. P., Manick, A., & Newman, R. (2020). Minimally invasive sampling of surface coatings  
1019 for protein identification by peptide mass fingerprinting: A case study with photographs.  
1020 *Journal of the American Institute for Conservation*, 59(3-4), 235–245.

1021 Kirby, D. P., Svoboda, M., Mazurek, J., Spaabæk, L. R., & Southon, J. (2023). Characterization  
1022 of an unusual coating on funerary portraits from Roman Egypt circa 100-300AD. *Heritage  
1023 Science*, 11(1).

1024 Knudsen, G. M., & Chalkley, R. J. (2011). The effect of using an inappropriate protein database  
1025 for proteomic data analysis. *PloS One*, 6(6), e20873.

1026 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017).  
1027 MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based  
1028 proteomics. *Nature Methods*, 14(5), 513–520.

1029 Lakens, D., Mesquida, C., Rasti, S., & Ditroilo, M. (2024). The benefits of preregistration and  
1030 Registered Reports. *Evidence-Based Toxicology*, 2(1).

1031 Lapatas, V., Stefanidakis, M., Jimenez, R. C., Via, A., & Schneider, M. V. (2015). Data  
1032 integration in biological research: an overview. *Journal of Biological Research*, 22(1), 9.

1033 Le Meillour, L., Sinet-Mathiot, V., Ásmundsdóttir, R. D., Hansen, J., Mylopotamitaki, D., Troché,  
1034 G., Xia, H., Herrera Bethencourt, J., Ruebens, K., Smith, G. M., Fagernäs, Z., & Welker, F.  
1035 (2024). Increasing sustainability in palaeoproteomics by optimizing digestion times for  
1036 large-scale archaeological bone analyses. *iScience*, 27(4), 109432.

1037 Li, H., Joh, Y. S., Kim, H., Paek, E., Lee, S.-W., & Hwang, K.-B. (2016). Evaluating the effect of  
1038 database inflation in proteogenomic search on sensitive and reliable peptide identification.  
1039 *BMC Genomics*, 17(Suppl 13), 1031.

1040 Lowenstein, J. M., Sarich, V. M., & Richardson, B. J. (1981). Albumin systematics of the extinct  
1041 mammoth and Tasmanian wolf. *Nature*, 291(5814), 409–411.

1042 Mackie, M., Rütther, P., Samodova, D., Di Gianvincenzo, F., Granzotto, C., Lyon, D., Peggie, D.  
1043 A., Howard, H., Harrison, L., Jensen, L. J., Olsen, J. V., & Cappellini, E. (2018).  
1044 Palaeoproteomic profiling of conservation layers on a 14th Century Italian wall painting.  
1045 *Angewandte Chemie*, 57(25), 7369–7374.

1046 Mann, A. E., Fellows Yates, J. A., Fagernäs, Z., Austin, R. M., Nelson, E. A., & Hofman, C. A.

1047 (2023). Do I have something in my teeth? The trouble with genetic analyses of diet from  
1048 archaeological dental calculus. *Quaternary International*, 653, 33–46.

1049 Miller, G. H., Beaumont, P. B., Jull, A. J., & Johnson, B. (1992). Pleistocene geochronology and  
1050 palaeothermometry from protein diagenesis in ostrich eggshells: implications for the  
1051 evolution of modern humans. *Philosophical Transactions of the Royal Society B: Biological  
1052 Sciences*, 337(1280), 149–157.

1053 Nair, B., Palomo, I. R., Markussen, B., Wiuf, C., Fiddyment, S., & Collins, M. J. (2023).  
1054 Parchment Glutamine Index (PQI): A novel method to estimate glutamine deamidation  
1055 levels in parchment collagen obtained from low-quality MALDI-TOF data. *Peer Community  
1056 Journal*, 3(e10), e10.

1057 Nel, T. H., Peters, C., Richter, K. K., Henshilwood, C., van Niekerk, K., & Douka, K. (2023).  
1058 Peptide mass fingerprinting as a tool to assess micromammal biodiversity in Pleistocene  
1059 South Africa: The case of Klipdrift Shelter. *Quaternary Science Reviews*, 322, 108380.

1060 Ottoni, C., Borić, D., Cheronet, O., Sparacello, V., Dori, I., Coppa, A., Antonović, D., Vujević, D.,  
1061 Douglas Price, T., Pinhasi, R., & Cristiani, E. (2021). Tracking the transition to agriculture in  
1062 Southern Europe through ancient DNA analysis of dental calculus. *Proceedings of the  
1063 National Academy of Sciences of the United States of America*, 118(32).

1064 Paleoanthropology Society & European Society for the study of Human Evolution. (2024). 14th  
1065 Annual Meeting of the European Society for the study of Human Evolution Abstracts  
1066 Zagreb, 11-15 September 2024. *PalaeoAnthropology*, 2, 365–558.

1067 Pálsdóttir, A. H., Bläuer, A., Rannamäe, E., Boessenkool, S., & Hallsson, J. H. (2019). Not a  
1068 limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains.  
1069 *Royal Society Open Science*, 6(10), 191059.

1070 Parker, R., Taylor, A., Peng, X., Nicastrì, A., Zerweck, J., Reimer, U., Wenschuh, H.,  
1071 Schnatbaum, K., & Ternette, N. (2021). The choice of search engine affects sequencing  
1072 depth and HLA class I allele-specific peptide repertoires. *Molecular & Cellular Proteomics*,  
1073 20, 100124.

1074 Paterson, R. S., Madupe, P. P., & Cappellini, E. (2024). Paleoproteomics sheds light on million-  
1075 year-old fossils. *Nature Reviews. Molecular Cell Biology*, 26(1), 1–2.

1076 Patramanis, I., Ramos-Madrigal, J., Cappellini, E., & Racimo, F. (2023). PaleoProPhyler: a  
1077 reproducible pipeline for phylogenetic inference using ancient proteins. *Peer Community  
1078 Journal*, 3.

1079 Penkman, K. E. H., Preece, R. C., Bridgland, D. R., Keen, D. H., Meijer, T., Parfitt, S. A., White,  
1080 T. S., & Collins, M. J. (2011). A chronological framework for the British Quaternary based  
1081 on Bithynia opercula. *Nature*, 476(7361), 446–449.

1082 Perez-Riverol, Y., Bittremieux, W., Noble, W. S., Martens, L., Bilbao, A., Lazear, M. R., Grüning,  
1083 B., Katz, D. S., MacCoss, M. J., Dai, C., Eng, J. K., Bouwmeester, R., Shortreed, M. R.,  
1084 Audain, E., Sachsenberg, T., Van Goey, J., Wallmann, G., Wen, B., Käll, L., & Fondrie, W.  
1085 E. (2024). Open-source and FAIR Research Software for Proteomics. *ChemRxiv*.

1086 Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein  
1087 identification by searching sequence databases using mass spectrometry data.  
1088 *Electrophoresis*, 20(18), 3551–3567.

1089 Pestle, W. J., Crowley, B. E., & Weirauch, M. T. (2014). Quantifying inter-laboratory variability in  
1090 stable isotope analysis of ancient skeletal remains. *PloS One*, 9(7), e102844.

1091 Peters, C., Richter, K. K., Manne, T., Dortch, J., Paterson, A., Travouillon, K., Louys, J., Price,  
1092 G. J., Petraglia, M., Crowther, A., & Boivin, N. (2021). Species identification of Australian  
1093 marsupials using collagen fingerprinting. *Royal Society Open Science*, 8(10), 211229.

1094 Peters, C., Wang, Y., Vakil, V., Cramb, J., Dortch, J., Hocknull, S., Lawrence, R., Manne, T.,  
1095 Monks, C., Rössner, G. E., Ryan, H., Siversson, M., Ziegler, T., Louys, J., Price, G. J.,  
1096 Boivin, N., & Collins, M. J. (2023). Bone collagen from subtropical Australia is preserved for  
1097 more than 50,000 years. *Communications Earth & Environment*, 4(1), 1–8.

1098 Poujois, P., & Le Meillour, L. (2023). Palaeoproteomics protocol - arid environment samples.  
1099 *Protocols.io*.

1100 Presslee, S., Penkman, K., Fischer, R., Richards-Slidel, E., Southon, J., Hospitaleche, C. A.,  
1101 Collins, M., & MacPhee, R. (2021). Assessment of different screening methods for selecting  
1102 palaeontological bone samples for peptide sequencing. *Journal of Proteomics*,  
1103 230(103986), 103986.

1104 Quye, A., & Strlič, M. (2019). Ethical sampling guidance. London: *The Institute of Conservation*  
1105 *(ICON)*.

1106 Ramsøe, A., Crispin, M., Mackie, M., McGrath, K., Fischer, R., Demarchi, B., Collins, M. J.,  
1107 Hendy, J., & Speller, C. (2021). Assessing the degradation of ancient milk proteins through  
1108 site-specific deamidation patterns. *Scientific Reports*, 11(1), 7795.

1109 Ramsøe, A., van Heekeren, V., Ponce, P., Fischer, R., Barnes, I., Speller, C., & Collins, M. J.  
1110 (2020). DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of  
1111 members of ancient proteomes. *Journal of Archaeological Science*, 115, 105080.

1112 Rayfield, E. J., Theodor, J. M., & Polly, P. D. (2020). Fossils from conflict zones and  
1113 reproducibility of fossil-based scientific data [Letter to Editors].

1114 R Core Team. (2023). R: A Language and Environment for Statistical Computing.  
1115 <https://www.R-project.org/>

1116 Research Data Alliance International Indigenous Data Sovereignty Interest Group. (2019).  
1117 CARE Principles. *GIDA-Global.org*.

1118 Richter, K. K., Codlin, M., Seabrook, M., & Warinner, C. (2022). A primer for ZooMS  
1119 applications in archaeology. *Proceedings of the National Academy of Sciences of the*  
1120 *United States of America*, 119(20), e2109323119.

1121 Rodriguez Palomo, I., Nair, B., Chiang, Y., Dekker, J., Dartigues, B., Mackie, M., Evans, M.,  
1122 Macleod, R., Olsen, J. V., & Collins, M. J. (2024). Benchmarking the identification of a  
1123 single degraded protein to explore optimal search strategies for ancient proteins. *Peer*  
1124 *Community Journal*, 4(e107), 2023.12.15.571577.

1125 Rusu, I., Paica, I., Vulpoi, A., Radu, C., Mircea, C., Dobrinescu, C., Bodolică, V., & Kelemen, B.  
1126 (2019). Dual DNA-protein extraction from human archeological remains. *Archaeological*  
1127 *and Anthropological Sciences*, 11(7), 3299–3307.

1128 Rütther, P. L., Husic, I. M., Bangsgaard, P., Gregersen, K. M., Pantmann, P., Carvalho, M.,  
1129 Godinho, R. M., Friedl, L., Cascalheira, J., Taurozzi, A. J., Jørkov, M. L. S., Benedetti, M.  
1130 M., Haws, J., Bicho, N., Welker, F., Cappellini, E., & Olsen, J. V. (2022). SPIN enables high  
1131 throughput species identification of archaeological bone by proteomics. *Nature*  
1132 *Communications*, 13(1), 2458.

1133 Saitta, E. T., Vinther, J., Crisp, M. K., Abbott, G. D., Wheeler, L., Presslee, S., Kaye, T. G., Bull,  
1134 I., Fletcher, I., Chen, X., Vidal, D., Sanguino, F., Buscalioni, Á. D., Calvo, J., Sereno, P. C.,  
1135 Baumgart, S. L., Pittman, M., Collins, M. J., Sakalauskaite, J., ... Penkman, K. E. H. (2024).  
1136 Non-avian dinosaur eggshell calcite can contain ancient, endogenous amino acids.  
1137 *Geochimica et Cosmochimica Acta*, 365, 1–20.

1138 Santucci, V. L., Newman, P. B., & Taff, B. D. (2016). Toward a conceptual framework for  
1139 assessing the human dimensions of paleontological resources. *Fossil Record* 5, *New*  
1140 *Mexico Museum of Natural History and Science Bulletin* 74, 239–248.

1141 Sawafuji, R., Cappellini, E., Nagaoka, T., Fotakis, A. K., Jersie-Christensen, R. R., Olsen, J. V.,  
1142 Hirata, K., & Ueda, S. (2017). Proteomic profiling of archaeological human bone. *Royal*  
1143 *Society Open Science*, 4(6), 161004.

1144 Schroeter, E. R., & Cleland, T. P. (2016). Glutamine deamidation: an indicator of antiquity, or  
1145 preservational quality? *Rapid Communications in Mass Spectrometry*, 30(2), 251–255.

1146 Schroeter, E. R., DeHart, C. J., Cleland, T. P., Zheng, W., Thomas, P. M., Kelleher, N. L., Bern,  
1147 M., & Schweitzer, M. H. (2017). Expansion for the *Brachylophosaurus canadensis* collagen  
1148 I sequence and additional evidence of the preservation of Cretaceous protein. *Journal of*

1149 *Proteome Research*, 16(2), 920–932.

1150 Scott, A., Power, R. C., Altmann-Wendling, V., Artzy, M., Martin, M. A. S., Eisenmann, S.,  
1151 Hagan, R., Salazar-García, D. C., Salmon, Y., Yegorov, D., Milevski, I., Finkelstein, I.,  
1152 Stockhammer, P. W., & Warinner, C. (2021). Exotic foods reveal contact between South  
1153 Asia and the Near East during the second millennium BCE. *Proceedings of the National  
1154 Academy of Sciences of the United States of America*, 118(2).

1155 Scott, A., Reinhold, S., Hermes, T., Kalmykov, A. A., Belinskiy, A., Buzhilova, A., Berezina, N.,  
1156 Kantorovich, A. R., Maslov, V. E., Guliyev, F., Lyonnet, B., Gasimov, P., Jalilov, B., Eminli,  
1157 J., Iskandarov, E., Hammer, E., Nugent, S. E., Hagan, R., Majander, K., ... Warinner, C.  
1158 (2022). Emergence and intensification of dairying in the Caucasus and Eurasian steppes.  
1159 *Nature Ecology & Evolution*, 6(6), 813–822.

1160 Scott, A., & Warinner, C. (2020). Ancient Proteins Extraction Protocol. *Protocols.io*.

1161 Sinet-Mathiot, V., Rendu, W., Steele, T. E., Spasov, R., Madelaine, S., Renou, S., Soulier, M.-  
1162 C., Martisius, N. L., Aldeias, V., Enderova, E., Goldberg, P., McPherron, S. J. P., Rezek, Z.,  
1163 Sandgathe, D., Sirakov, N., Sirakova, S., Soressi, M., Tsanova, T., Turq, A., ... Smith, G.  
1164 M. (2023). Identifying the unidentified fauna enhances insights into hominin subsistence  
1165 strategies during the Middle to Upper Palaeolithic transition. *Archaeological and  
1166 Anthropological Sciences*, 15(9), 1–18.

1167 Sinet-Mathiot, V., Smith, G. M., Romandini, M., Wilcke, A., Peresani, M., Hublin, J.-J., & Welker,  
1168 F. (2019). Combining ZooMS and zooarchaeology to study Late Pleistocene hominin  
1169 behaviour at Fumane (Italy). *Scientific Reports*, 9(1), 12350.

1170 Society of Vertebrate Paleontology. (2025). Member Bylaw on Ethics Statement.

1171 Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S.,  
1172 Esterling, K. M., & Nosek, B. A. (2021). Initial evidence of research quality of registered  
1173 reports compared with the standard publishing model. *Nature Human Behaviour*, 5(8), 990–  
1174 997.

1175 Solazzo, C., & Phipps, E. (2022). Chasing the elusive viscacha in Precolumbian textiles at the  
1176 intersection of art and science. *Journal of Archaeological Science*, 140, 105575.

1177 Solazzo, C., Wilson, J., Dyer, J. M., Clerens, S., Plowman, J. E., von Holstein, I., Walton  
1178 Rogers, P., Peacock, E. E., & Collins, M. J. (2014). Modeling deamidation in sheep  $\alpha$ -  
1179 keratin peptides and application to archeological wool textiles. *Analytical Chemistry*, 86(1),  
1180 567–575.

1181 Solntsev, S. K., Shortreed, M. R., Frey, B. L., & Smith, L. M. (2018). Enhanced global post-  
1182 translational modification discovery with MetaMorpheus. *Journal of Proteome Research*,  
1183 17(5), 1844–1851.

1184 Somel, M., Altınışık, N. E., Özer, F., & Ávila-Arcos, M. C. (2021). Collaborate equitably in  
1185 ancient DNA research and beyond. *Nature*, 600(7887), 37.

1186 Speller, C., van den Hurk, Y., Charpentier, A., Rodrigues, A., Gardeisen, A., Wilkens, B.,  
1187 McGrath, K., Rowsell, K., Spindler, L., Collins, M., & Hofreiter, M. (2016). Barcoding the  
1188 largest animals on Earth: ongoing challenges and molecular solutions in the taxonomic  
1189 identification of ancient cetaceans. *Philosophical Transactions of the Royal Society B:  
1190 Biological Sciences*, 371(1702).

1191 Standvoss, K., Kazezian, V., Lewke, B. R., Bastian, K., Chidambaram, S., Arafat, S., Alsharif,  
1192 U., Herrera-Melendez, A., Knipper, A.-D., Seco, B. M. S., Soto, N. N., Rakitzis, O.,  
1193 Steinecker, I., van Kronenberg Till, P., Zarebidaki, F., Abbasi, P., & Weissgerber, T. L.  
1194 (2024). Shortcut citations in the methods section: Frequency, problems, and strategies for  
1195 responsible reuse. *PLoS Biology*, 22(4), e3002562.

1196 Stewart, N. A., Gerlach, R. F., Gowland, R. L., Gron, K. J., & Montgomery, J. (2017). Sex  
1197 determination of human remains from peptides in tooth enamel. *Proceedings of the  
1198 National Academy of Sciences of the United States of America*, 114(52), 13649–13654.

1199 Sticker, A., Martens, L., & Clement, L. (2017). Mass spectrometrists should search for all

1200 peptides, but assess only the ones they care about. *Nature Methods*, 14(7), 643–644.

1201 Strohaln, M., Kavan, D., Novák, P., Volný, M., & Havlíček, V. (2010). mMass 3: a cross-

1202 platform software environment for precise analysis of mass spectrometric data. *Analytical*

1203 *Chemistry*, 82(11), 4648–4651.

1204 Taylor, G. K., & Goodlett, D. R. (2005). Rules governing protein identification by mass

1205 spectrometry. *Rapid Communications in Mass Spectrometry*, 19(23), 3420.

1206 Teasdale, M. D., Fiddymment, S., Vnouček, J., Mattiangeli, V., Speller, C., Binois, A., Carver, M.,

1207 Dand, C., Newfield, T. P., Webb, C. C., Bradley, D. G., & Collins, M. J. (2017). The York

1208 Gospels: a 1000-year biological palimpsest. *Royal Society Open Science*, 4(10), 170988.

1209 Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie,

1210 A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A.,

1211 Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., ... Hassabis, D. (2021). Highly

1212 accurate protein structure prediction for the human proteome. *Nature*, 596(7873), 590–596.

1213 University of York. (2025). Palaeoproteomics and ZooMS. University of York.

1214 [https://www.york.ac.uk/archaeology/research/research-centres/bioarch/research-](https://www.york.ac.uk/archaeology/research/research-centres/bioarch/research-themes/palaeoproteomics/)

1215 [themes/palaeoproteomics/](https://www.york.ac.uk/archaeology/research/research-centres/bioarch/research-themes/palaeoproteomics/)

1216 Van Den Bossche, T., Kunath, B. J., Schallert, K., Schäpe, S. S., Abraham, P. E., Armengaud,

1217 J., Arntzen, M. Ø., Bassignani, A., Benndorf, D., Fuchs, S., Giannone, R. J., Griffin, T. J.,

1218 Hagen, L. H., Halder, R., Henry, C., Hettich, R. L., Heyer, R., Jagtap, P., Jehmlich, N., ...

1219 Muth, T. (2021). Critical Assessment of MetaProteome Investigation (CAMPI): a multi-

1220 laboratory comparison of established workflows. *Nature Communications*, 12(1), 7305.

1221 van der Sluis, L. G., McGrath, K., Thil, F., Cersoy, S., Pétilion, J.-M., & Zazzo, A. (2023).

1222 Identification and tentative removal of collagen glue in Palaeolithic worked bone objects:

1223 implications for ZooMS and radiocarbon dating. *Scientific Reports*, 13(1), 22119.

1224 Végh, E. I., & Douka, K. (2024). SpecieScan: Semi-Automated taxonomic identification of bone

1225 collagen peptides from MALDI-ToF MS spectra. *Bioinformatics*, 40(3), btae054.

1226 Velsko, I. M., Overmyer, K. A., Speller, C., Klaus, L., Collins, M. J., Loe, L., Frantz, L. A. F.,

1227 Sankaranarayanan, K., Lewis, C. M., Jr, Martinez, J. B. R., Chaves, E., Coon, J. J., Larson,

1228 G., & Warinner, C. (2017). The dental calculus metabolome in modern and historic

1229 samples. *Metabolomics: Official Journal of the Metabolomic Society*, 13(11), 134.

1230 Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., & Yates, J. R. (2004). Automated

1231 approach for quantitative analysis of complex peptide mixtures from tandem mass spectra.

1232 *Nature Methods*, 1(1), 39–45.

1233 Viñas-Caron, L. C., & Brandt, L. Ø. (2025). Clothomics: a practical guide to understand the

1234 opportunities and challenges of omics-based methods in archaeological cloth research.

1235 *NPJ Heritage Science*, 13(1), 80.

1236 Viñas-Caron, L. C., Rodríguez Palomo, I., Fazlic, N., Vnouček, J., Driscoll, M., Fiddymment, S., &

1237 Collins, M. J. (2023). A biological reading of a palimpsest. *iScience*, 26(6), 106786.

1238 Warinner, C., Hendy, J., Speller, C., Cappellini, E., Fischer, R., Trachsel, C., Arneborg, J.,

1239 Lynnerup, N., Craig, O. E., Swallow, D. M., Fotakis, A., Christensen, R. J., Olsen, J. V.,

1240 Liebert, A., Montalva, N., Fiddymment, S., Charlton, S., Mackie, M., Canci, A., ... Collins, M.

1241 J. (2014). Direct evidence of milk consumption from ancient human dental calculus.

1242 *Scientific Reports*, 4, 7104.

1243 Warinner, C., Korzow Richter, K., & Collins, M. J. (2022). Paleoproteomics. *Chemical Reviews*,

1244 122(16), 13401–13446.

1245 Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A.,

1246 Hancock, Y., Tito, R. Y., Fiddymment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U.,

1247 Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., ...

1248 Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity.

1249 *Nature Genetics*, 46(4), 336–344.

1250 Wehmiller, J. F., Hare, P. E., & Kujala, G. A. (1976). Amino acids in fossil corals: racemization

1251 (epimerization) reactions and their implications for diagenetic models and geochronological  
1252 studies. *Geochimica et Cosmochimica Acta*, 40(7), 763–776.

1253 Welker, F., Collins, M. J., Thomas, J. A., Wadsley, M., Brace, S., Cappellini, E., Turvey, S. T.,  
1254 Reguero, M., Gelfo, J. N., Kramarz, A., Burger, J., Thomas-Oates, J., Ashford, D. A.,  
1255 Ashton, P. D., Rowsell, K., Porter, D. M., Kessler, B., Fischer, R., Baessmann, C., ...  
1256 MacPhee, R. D. E. (2015). Ancient proteins resolve the evolutionary history of Darwin's  
1257 South American ungulates. *Nature*, 522(7554), 81–84.

1258 Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer,  
1259 M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B. M., Stewart, J.  
1260 R., Pääbo, S., Collins, M. J., & Hublin, J.-J. (2016). Palaeoproteomic evidence identifies  
1261 archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings*  
1262 *of the National Academy of Sciences of the United States of America*, 113(40), 11162–  
1263 11167.

1264 Welker, F., Ramos-Madrugal, J., Gutenbrunner, P., Mackie, M., Tiwary, S., Rakownikow Jersie-  
1265 Christensen, R., Chiva, C., Dickinson, M. R., Kuhlwilm, M., de Manuel, M., Gelabert, P.,  
1266 Martínón-Torres, M., Margvelashvili, A., Arsuaga, J. L., Carbonell, E., Marques-Bonet, T.,  
1267 Penkman, K., Sabidó, E., Cox, J., ... Cappellini, E. (2020). The dental proteome of *Homo*  
1268 *antecessor*. *Nature*, 580(7802), 235–238.

1269 Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

1270 Wilkin, S., Lanigan, L. T., Montes, N., Sharma, M., Avanzi, C., Sejdiu, D., Majander, K.,  
1271 Pfrengle, S., Chiang, Y., Kunz, L., Dittmann, A., Rühli, F., Singh, P., Coll, M. F., Collins, M.  
1272 J., Taurozzi, A. J., & Schuenemann, V. J. (2024). Sequential trypsin and ProAlanase  
1273 digestions unearth immunological protein biomarkers shrouded by skeletal collagen.  
1274 *iScience*, 27(5), 109663.

1275 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A.,  
1276 Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes,  
1277 A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ...  
1278 Mons, B. (2016). The FAIR Guiding Principles for scientific data management and  
1279 stewardship. *Scientific Data*, 3(1), 160018.

1280 Wilkin, S., Ventresca Miller, A., Fernandes, R., Spengler, R., Taylor, W. T.-T., Brown, D. R.,  
1281 Reich, D., Kennett, D. J., Culleton, B. J., Kunz, L., Fortes, C., Kitova, A., Kuznetsov, P.,  
1282 Epimakhov, A., Zaibert, V. F., Outram, A. K., Kitov, E., Khokhlov, A., Anthony, D., & Boivin,  
1283 N. (2021). Dairying enabled Early Bronze Age Yamnaya steppe expansions. *Nature*,  
1284 598(7882), 629–633.

1285 Wilson, J., van Doorn, N. L., & Collins, M. J. (2012). Assessing the extent of bone degradation  
1286 using glutamine deamidation in collagen. *Analytical Chemistry*, 84(21), 9041–9048.

1287 Winter, R. M., de Kock, W., Mackie, M., Ramsøe, M., Desiderà, E., Collins, M., Guidetti, P.,  
1288 Presslee, S., Alegre, M. M., Oueslati, T., Muniz, A. M., Michailidis, D., van den Hurk, Y.,  
1289 Taurozzi, A. J., & Çakırlar, C. (2023). Grouping groupers in the Mediterranean: Ecological  
1290 baselines revealed by ancient proteins. *Ecology and Evolution*, 13(10), e10625.

1291 Xia, H., Zhang, D., Wang, J., Fagernäs, Z., Li, T., Li, Y., Yao, J., Lin, D., Troché, G., Smith, G.  
1292 M., Chen, X., Cheng, T., Shen, X., Han, Y., Olsen, J. V., Shen, Z., Pei, Z., Hublin, J.-J.,  
1293 Chen, F., & Welker, F. (2024). Middle and Late Pleistocene Denisovan subsistence at  
1294 Baishiya Karst Cave. *Nature*, 1–6.

1295 Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., &  
1296 Ma, B. (2012). PEAKS DB: de novo sequencing assisted database search for sensitive and  
1297 accurate peptide identification. *Molecular & Cellular Proteomics*, 11(4), M111.010587.