

Burnt to a crisp, Disentangling the taxonomic composition of foodcrusts via palaeoproteomics

Joannes Adrianus Antonius Dekker

Double PhD degree

Host university:
University of York
Archaeology

Partner university:
University of Copenhagen
Faculty of Health and Medical Sciences

November 2024

English abstract

The main application of palaeoproteomics is taxonomic identification. The most commonly analysed sample types in palaeoproteomics are osseous materials. A shared characteristic of osseous objects is that they each derive from a single species, allowing researchers to focus on the peptides with the highest taxonomic resolution. However, other material types, such as charred organic residues (foodcrusts), are gaining attention in particular due to their potential to study past diets. The analysis of foodcrusts has its own advantages and disadvantages. On the one hand foodcrusts provide direct evidence of cooking, facilitating the investigation of culinary behaviour rather than just diet. On the other hand, apart from severe heating, which damages and modifies the remaining proteins, it is not clear what is required for the formation of foodcrusts. Additionally, no *a priori* assumptions can be made regarding the contents of a foodcrust, of which the taxonomic diversity can easily stretch over several taxonomic kingdoms.

This thesis aims to contribute to the analysis of foodcrusts with several methodological experiments and test what information foodcrusts can provide by applying it to an archaeological case study. First of all, to highlight the taxonomic resolution of palaeoproteomics in a controlled setting, we apply several proteomic workflows to a bone tool from a Dutch Bronze Age site. Secondly, to investigate the formation of foodcrusts and the degree to which the recovered proteins represent the cooked contents, a set of foodcrusts with known composition was generated and analysed. Another experiment consisted of a blind test of identifying the resources cooked in experimental foodcrusts, highlighting both the over- and underrepresented taxa in proteomic analysis. Finally, foodcrusts from the Mesolithic-Neolithic site Syltholm II (MLF00906-I/II) were analysed with palaeoproteomics, lipid analysis, and SEM to highlight the different resources identified by each of these methods.

Dansk resumé

I oldproteomik er hovedanvendelsen artsbestemmelse, hvor de mest almindelige analyserede prøveekstrakter er knogleobjekter. Et delt karakteristikum mellem knogleobjekter er, at de hver især stammer fra bestemte arter hvilket muliggør for forskere at fokusere på peptider med den største taksonomiske opløsning. Midlertidigt er der andre typer materialer, såsom forkullede organiske rester (madskorper), der er ved at vinde opmærksomhed grundet deres potentiale til at studere forhistorisk kost. Studiet af madskorper har både sine fordele og ulemper. På den ene side er madskorper et direkte bevis for madlavning, der faciliterer undersøgelsen af kulinariske adfærdsmønstre snarere end den foreliggende kost. På den anden side, udover ekstrem varmepåførelse som skader og modificerer de eksisterende proteiner, er det ikke tydeligt, hvad der kræves i tilblivelsen af madskorper. Endvidere kan der ikke på forhånd laves antagelser om indholdet af madskorper, da den taksonomiske diversitet let kan strække sig over flere taksonomiske riger.

Denne afhandling forsøger at bidrage til analysen af madskorper ved hjælp af forskellige metodologiske eksperimenter samt teste, hvilke informationer madskorper kan indeholde i undersøgelsen af konkrete forhistoriske eksempler. Først fremhæves den taksonomiske opløsning gennem oldproteomik i et kontrolleret miljø, hvor forskellige proteomikarbejdsforløb anvendes på bengenstande fra en nederlandsk bronzealderboplads. Dernæst undersøges formationsprocesserne omkring madskorper og graden af mulige udvundne proteiner i kontekster, hvor mad er tilberedt. I den forbindelse blev en række madskorper med kendte proteinsammensætninger genereret og analyseret. Et andet eksperiment bestod af en blindtest, hvor de tilberedte komponenter blev identificeret gennem madskorperne, hvilket fremhæver over- og underrepræsenterede taxa i proteomikanalyser. Til sidst blev madskorper fra den mesolitiske og neolitiske fundplads Syltholm II (MLF00906-I/II) analyseret gennem oldproteomik, lipidanalyse og SEM for at fremhæve de forskellige ressourcer, der blev identificeret gennem disse metoder.

List of contents

English abstract	2
Dansk resumé	3
List of contents	4
Figures in thesis	7
Figures	7
Tables	8
List of Accompanying Material	10
Supplementary Information	10
Signed declarations of authorship	10
Acknowledgements	11
Statement of authorship	13
Abbreviations	14
1 Introduction	16
1.1 The merits of foodcrust	17
1.2 Analysing foodcrust	18
1.2.1 Scanning electron microscopy (SEM)	18
1.2.2 Lipid analysis	19
1.2.3 Palaeoproteomics	20
1.3 The archaeological relevance of foodcrust	26
1.4 Objectives	28
1.5 Thesis outline	30
Chapter 2:	
Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands	30
Chapter 3:	
An unclear slate, discrepancies between food input and recovered lipid and protein signal from experimental foodcrusts.	30
Chapter 4:	
Trial by fire, a blind test of the proteomic analysis of experimental foodcrust	31
Chapter 5:	
Palaeoproteomic, lipid and SEM analysis of foodcrusts from Syltholm II (MLF00906-I/II), Denmark, reveals dairy consumption in the Mesolithic Ertebølle	31
Chapter 6: Discussion	31
Chapter 7: Conclusion	32
References	33
2 Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands	47
2.1 Introduction	49
2.2 Materials and methods	50
2.2.1 Description of the site and context	50
2.2.2 Description of the artefact (Specimen 2199)	51
2.2.3 Use-wear methods	52
2.2.4 ZooMS	52
2.2.5 LC-MS/MS data acquisition for SPIN data analysis	53
2.2.6 MaxQuant data analysis	54
2.2.7 Bioinformatics data analysis	54
2.3 Results	55
2.3.1 Use-wear results	55
2.3.2 ZooMS	56
2.3.3 SPIN data analysis	59

2.3.4 Selected baleen whale database search	59
2.4 Discussion	64
2.4.1 Contextualising molecular taxonomic identification	64
2.4.2 Bronze Age whale exploitation: hunting or scavenging?	66
2.4.3 Contextualising the tool in the Bronze Age	67
2.5 Conclusion	68
Acknowledgements	68
Funding	69
References	69
3 An unclear slate, discrepancies between food input and recovered protein signal from experimental foodcrusts.	76
CRedit statement:	76
Acknowledgements:	76
Funding	77
Abstract	77
3.1 Background	77
3.2 Materials	79
3.3 Methods	80
3.3.1 Cooking experimental protocol	80
3.3.2 Protein extraction	82
3.3.3 Protein Mass Spectrometry analysis	82
3.3.4 Protein data analysis	83
3.3.5 Foodcrust quantification	84
3.4 Results	84
3.5 Discussion	90
3.5.1 Protein recovery	90
3.5.2 Disentangling mixtures & taxonomic representativeness	91
3.5.3 Foodcrust abundance	94
3.6 Conclusion	95
References	96
4 Trial by fire, a blind test of the proteomic analysis of experimental foodcrust	102
CRedit statement	102
Acknowledgements:	102
Funding statement:	103
Abstract	103
4.1 Introduction	103
4.2 Materials	105
4.3 Methods	108
4.3.1 Protein extraction	108
4.3.2 Mass spectrometry data acquisition	109
4.3.3 Bioinformatic analysis	111
4.4 Results	111
4.4.1 Dataset 2	114
4.4.2 Dataset 1	115
4.5 Discussion	121
4.5.1 The elephant in the room	121
4.5.2 Chaotic mixture	124
4.5.3 Attempted methodological improvements	124
4.5.4 Returning to order	125
4.5.5 Implications for archaeological research	126

4.6 Conclusion	127
References	128
5 Palaeoproteomic, lipid and SEM analysis of foodcrusts from Syltholm II (MLF00906-I/II), Denmark, reveals dairy consumption in the Mesolithic Ertebølle	133
CRediT statement	133
Acknowledgements and funding sources	134
Abstract	134
Significance statement	134
5.1 Introduction	135
5.2 Results	137
5.2.1 Radiocarbon dating	137
5.2.2 SEM	138
5.2.3 Lipid analysis	140
5.2.4 Proteomics	140
5.3 Discussion	143
5.3.1 Reliability	143
5.3.2 Taxonomic accuracy	144
5.3.3 Combining the data	144
5.3.4 Dairy in the Mesolithic	145
5.3.5 Change and continuity	145
5.4 Materials and methods	146
5.4.1 Ceramic typo-technological analysis	146
5.4.2 Radiocarbon analysis	147
5.4.3 Foodcrust microscopic analysis	147
5.4.4 Lipid analysis	147
5.4.5 Proteomic analysis	147
References	150
6 General discussion	156
6.1 The nature of foodcrust	158
6.1.1 A different beast by a similar name	159
6.1.2 Chemical composition	161
6.1.3 Fair representation	162
6.2 Bottom up or top down	163
6.3 Bioinformatics	163
6.3.1 PSMs	164
6.3.2 Databases	165
6.3.3 Taxonomic validation	167
6.3.4 A solution? Sequential searching	168
6.4 The place of proteomics in the archaeology of past diet	169
6.4.1 Tissue specificity	169
6.4.2 Taxonomic identifications	170
6.4.3 Reevaluating relevance	172
6.5 Future directions	173
References	175
7 Conclusion	184

Figures in thesis

Figures

Chapter	Figure number	Figure title	Page number
2	1	Map of the Netherlands with a star indicating the site location.	49
2	2	Photograph of the bone artefact, 2199, and drawing show the cross sections of the tool.	50
2	3	Photos of the use-wear on the bone object.	54
2	4	ZooMS spectra.	55
2	5	Highest scoring MS2 spectrum of the COL1A1 918-957 <i>Eubalaena</i> sp. Peptide covering the 935 SAP.	59
2	6	Highest scoring MS2 spectrum of the COL1A1 1170-1213 peptide covering the 1193 SAP.	60
2	7	Highest scoring MS2 spectrum of the COL1A2 262-280 peptide covering the 276 SAP.	61
2	8	Baleen whale phylogeny and presence in relevant proteomic databases.	63
3	1	Overview of the experimental conditions.	80
3	2	Foodcrust formed after the ninth cooking event.	81
3	3	Number of proteins identified in each experimental condition after cooking event 1 (A), cooking event 5 (B), cooking event 9 (C) and post-excavation (D).	85
3	4	Distribution of the number of proteins identified per sampling point.	86
3	5	Distribution of the intensity of the three most intense peptides per protein.	87
3	6	Tissue associations of the identified proteins per species.	89

3	7	Volume of foodcrust accumulated after nine cooking events, sorted by experimental conditions listed in Fig. 1.	90
4	1	Overview of the different methods used in the analysis of the samples.	105
4	2	Number of proteins (triangles) and peptides (circles) per sample.	113
5	1	Number of samples and their cultural associations per method.	137
5	2	95% confidence intervals of calibrated ¹⁴ C dates of five EBK and 5 TRB foodcrusts using OxCal v4.4.4 (Bronk Ramsey 2020).	138
5	3	SEM images of foodcrust.	139
5	4	Compound-specific isotope ratios.	140
5	5	The number of peptides and associated taxa per sample after validation.	142
6	1	Examples of calcified deposits and “bone powder” residues.	160

Tables

Chapter	Table number	Table title	Page number
2	1	Observed biomarkers and matching taxa.	56
2	2	SPIN taxonomic identifications of the bone tool (2199).	57
2	3	Number of peptides per species for the most abundant protein group for COL1A1 and COL1A2	58
2	4	SAPs differing between <i>Eubalaena</i> sp. And <i>Balaena mysticetus</i> , and the number of matching observed PSMs.	58
3	1	Identified proteins belonging to taxa not included in the samples's mixture.	85

3	2	Uniprot accession number, full name and associated species of the proteins found in at least one rim sample from each sampling point.	88
3	3	Revised taxonomic identifications using BLAST of potentially intrusive proteins	92
3	4	Recovery rate of each species per cooking event.	93
4	1	Overview of the analysed samples.	106
4	2	List of analysed experimental foodcrusts and their original contents.	107
4	3	Initial interpretation of proteomic results versus the cooked foods (CF) for each foodcrust.	115
4	4	Count of distinct peptide sequences per taxonomic group for all samples in Dataset 1.	117
4	5	Count of distinct peptide sequences per taxonomic group for all samples in Dataset 2.	117
4	6	Initial interpretation of the proteomic results of dataset 1 and the cooked foods (CF) for each experimental foodcrust.	120
4	7	BLAST taxonomic assignment of ProteoClade's species specific peptides.	122
4	8	Sequence taxonomic identifications of dataset 1 grouped at order level with a minimum six sequences.	126

List of Accompanying Material

Supplementary Information

File	Description	File format
Chapter 3 SI 1	Nutritional composition of the used foods	csv
Chapter 3 SI 2.1	Plant find list	csv
Chapter 3 SI 2.2	Animal find list	csv
Chapter 3 SI 2.3	Plant species list	csv
Chapter 3 SI 2.4	Fish species list	csv
Chapter 3 SI 2.5	Mammal species list	csv
Chapter 3 SI 2.6	Bird species list	csv
Chapter 3 SI 2.7	Shell species list	csv
Chapter 3 SI 3	Metamorpheus metadata	zip
Chapter 3 SI 4	List of identified proteins	csv
Chapter 3 SI 5	Overview of the number of the number of foodcrusts per cooking event a protein was identified in	csv
Chapter 5 SI 1	Radiocarbon result rapport	csv
Chapter 5 SI 2.1	All identified peptide sequences meeting initial identification criteria	csv
Chapter 5 SI 2.2	BLAST results	csv
Chapter 5 SI 2.3	Anubis deamidation calculations	csv
Chapter 5 SI 3	Overview of performed analyses per sample	csv

Signed declarations of authorship

KU Declaration of co-authorship for Chapter 2,
KU Declaration of co-authorship for Chapter 3,
KU Declaration of co-authorship for Chapter 4,
KU Declaration of co-authorship for Chapter 5,
York Declaration of co-authorship for Chapter 3,
York Declaration of co-authorship for Chapter 4,
York Declaration of co-authorship for Chapter 5.

Acknowledgements

To start I would like to thank my supervisors. Jessie, I would like to thank you for helping not to get too distracted by an endless number of side projects and for giving me the freedom to pursue my ideas, even when they seemed less promising at first. Matthew, thank you for your unending enthusiasm, it is truly inspiring. I have learned so much from you both and I am very grateful for your help in completing this journey.

Additionally, Frido, I would like to thank you for helping me feel welcome in Copenhagen when I arrived, as well as your advice. I am really glad with how the whale paper turned out and I greatly appreciate your role in getting that done.

I am also incredibly grateful to all my colleagues in York and Copenhagen. Whether it was during coffee breaks, conferences, escape rooms or board game evenings, your company was the best part of my stay in both cities. Of course, I should also thank my fellow ChemArch PhDs. You all made the long days at the workshops enjoyable and I have learned a lot from our collaborations.

From York, I would like to give special thanks to Alice, Mackenzie and Christina. We probably chatted too much during coffee breaks, but this thesis is a testament to the fact that some work got done. I really enjoyed our adventures, be they in Northumbria, Wales or the York moors.

Also not to forget my bouldering buddies: Daan, Ivy, Tabea and Sterre and Kasper. You guys literally helped me to reach the top (even though it was only three meters high) and perhaps more importantly to maintain some semblance of a work-life balance.

From Copenhagen, I would like to thank Diljá, Laura, Leire, Liam, Louise, Viridiana and Zandra. Thank you for making the office such an enjoyable environment and for indulging me in our trips to Bastard Café. Of course I could not forget Jakob, it was wonderful to work together and I enjoyed braving the Swedish snow together. It's a shame the French hike did not work out, but maybe we can still go somewhere our beards won't freeze.

Max and Yun, thank you so much for being as obsessed with board games as me. I will fondly look back at days and evenings spent peering over obscure rule books and devising cunning master plans only to be foiled by the roll of the dice or your own devious stratagems.

I also want to give special thanks to the amazing support staff; technicians, administrators and all others, that I have worked with at both the University of York and University of Copenhagen. Without all these wonderful people much of my work and that of others would have either failed, been delayed or been impossible.

In particular I would like to thank Chloë and Adam from the Centre of Excellence for Mass Spectrometry in York. As a mere archaeologist I am still perpetually amazed at the wonders of mass spectrometry and I am ever grateful for your guidance and support in making sure I actually got data out of my samples.

Since 2021, even before the start of this PhD, I have moved countries every year. This semi-nomadic lifestyle has been somewhat of a source of confusion for all my friends and family back home and, to a certain extent, myself. Which is why I would like to thank all

my friends back home in the Netherlands for still making it feel like home whenever I returned from my wanderings. Only being there a few times every year is definitely a stress test of friendships and I am incredibly grateful for you all sticking with me. I have not and will not take it for granted.

Hetzelfde geldt natuurlijk ook voor mijn familie. Mam, pap, Anna en Wijn, sorry voor het chaotische heen-en-weer, het missen van alle verjaardagen, familie-uitjes en een algemeen gebrek aan aanwezigheid, maar enorm bedankt voor alle steun, zowel moreel als praktisch. Ik heb geen idee hoe ik al dat verhuizen alleen voor elkaar zou hebben gekregen of hoe ik überhaupt dit proefschrift had kunnen doorstaan zonder jullie steun.

Oma, erg bedankt voor al uw kaarten! Het was erg fijn om iets kleurrijks te hebben in al die kale kamers en het was leuk om elke week nieuws uit Nederland uit te wisselen!

Statement of authorship

This thesis is an original work written solely for the purpose of fulfilling the requirements of a double doctoral degree for both the University of York and the University of Copenhagen. The contents of this thesis have not previously been presented for a degree or qualification at these two universities or elsewhere. All sources used for the writing of this thesis are cited and acknowledged in the references.

Chapters 1, 6 and 7 are solely written by the author, Joannes Dekker. Chapters 2, 3, 4 and 5 are manuscripts containing contributions from multiple authors with the main contributor being Joannes Dekker. Of these chapters, Chapter 2 has been published in the *Peer Community Journal*. Chapter 3 has been submitted to the journal *Royal Society Open Science* and is currently under consideration. Chapters 4 and 5 are manuscripts written with the intent for submission. Each manuscript chapter starts with a CRediT statement detailing the contributions of all involved authors.

Abbreviations

AA	Acrylamide
ACN	Acetonitrile
AGC	Automatic gain control
AGE	Advanced Glycation Endproduct
AmBic	Ammonium bicarbonate
AMS	Accelerator Mass Spectrometer
APAA	ω -(o-alkylphenyl) alkanolic acids
ATR-FTIR	Attenuated Total Reflection - Fourier Transform Infrared
BLG	Beta-lactoglobulin
BME	β -mercaptoethanol
CAA	Chloroacetamide
CHCA	α -cyano-4-hydroxycinnamic acid
CID	Collision-Induced Dissociation
DDA	Data Dependent Acquisition
DES	Deep Eutectic Solvent
DIA	Data Independent Acquisition
DTT	Dithiothreitol
EBK	Ertebølle culture
ECD	Electron Capture Dissociation
ESI	Electron Spray Ionisation
ETD	Electron Transfer Dissociation
FA	Formic acid
FASP	Filter-aided separation protocols
FDR	False Discovery Rate
FWHH	Full Width at Half Height
GASP	Gel-aided separation protocols
GC-C-IRMS	Gas Chromatography - Combustion - Isotope Ratio Mass Spectrometry
GC-MS	Gas Chromatography - Mass Spectrometry
GuHCl	Guanidine Hydrochloride
HCD	Higher-energy Collision Dissociation
HCl	Hydrochloric acid
IAA	Iodoacetamide

IAC	Iodoacetic acid
LC-MS/MS	Liquid Chromatography - tandem Mass Spectrometry
M-PER	Mammalian Protein Extraction Reagent
<i>m/z</i>	Mass to charge ratio
MALDI-TOF-MS	Matrix Assisted Laser Desorption/Ionisation - Time Of Flight - Mass Spectrometer
MRCA	Most Recent Common Ancestor
MRE	Marine Reservoir Effect
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NCE	Normalised Collision Energy
PMF	Peptide Mass Fingerprinting
PSM	Peptide Spectrum Match
PTM	Post Translational Modifications
SDS	Sodium dodecyl sulfate
SEM	Scanning Electron Microscopy
SAP	Single Amino acid Polymorphism
SP3	Single-Pot Solid-Phase-enhanced Sample Preparation
SPD	Samples per day
SPIN	Species by Proteome INvestigation
TCEP	Tris-(2-carboxyethyl)-phosphine
TFA	Trifluoroacetic acid
TRB	Funnelbeaker culture
UTR	UnTranslated Region
ZooMS	Zooarchaeology by Mass Spectrometry

1 Introduction

At its core, taxonomic identification is a fundamental part of the majority of palaeoproteomic applications. All proteomic research involves the assignment of an amino acid sequence to an ion based on its measured mass to charge ratio of and, in the case of tandem mass spectrometry, that of its fragments. The amino acid sequence is irrevocably linked to both the protein identity and its taxonomic identity. Hence, when studying a specimen with unknown taxonomic origin, protein identification necessarily involves taxonomic identification. Moreover, the taxonomic identification often plays a central role in the knowledge that palaeoproteomic research produces of the past. Palaeoproteomics has been applied to bone fragments in order to study what species were present at a site (Buckley et al. 2009; Buckley and Kansa 2011), it has been used to investigate how different taxa are related to each other (Welker et al. 2015), what species were used for the production of artefacts (Desmond et al. 2018; Bradfield et al. 2019; McGrath et al. 2019) and to study the conservation histories of historic paintings (Mackie et al. 2018). Although the aforementioned studies cover a wide range of topics, their research questions all revolve around the taxonomic identification of proteins.

Most of the commonly analysed materials in archaeology, such as bone, enamel and skin products are derived from a single species. Single-species objects greatly facilitate palaeoproteomic analysis in two ways. First of all, one can focus on the most taxonomically specific protein to assign a taxon to the sample. Secondly, one can combine the taxonomic information from multiple proteins to further specify the taxonomic identification. For these types of single-taxon objects it is often possible to distinguish between very closely related taxa, even up to the species level in some cases (Chen et al. 2019; R  ther et al. 2022).

In contrast to these single-taxon objects, there are a number of other materials, such as dental calculus, calcified deposits and foodcrusts that are composed of any number of different taxa and tissues. It is not possible to establish a priori how many taxa are present in these types of materials and as a result, the taxon of each protein recovered from these materials must be determined and interpreted individually.

Of the three, dental calculus, calcified deposits and foodcrusts, foodcrusts form a particularly challenging material. Foodcrusts are charred organic deposits encrusted on the wall of a container, usually ceramic (Heron and Craig 2015). There is some inconsistency in the terminology used to describe foodcrusts in the literature. The same type of residue has also been referred to as “chars”, “carbonised residues” (Heron and Craig 2015) or simply “organic residues” (Kubiak-Martens et al. 2024). This can create some confusion as organic residue analysis for example is also known to include the study of lipid residues preserved inside the ceramic matrix. The processes leading to the entrapment of biomolecules within ceramic and their subsequent survival (Craig and Collins 2002; Barker et al. 2012, 2018) are fundamentally different from the processes that lead to the formation of residues on top of a container. These differences likely impact the process required to extract the proteins from the material therefore it is important to maintain a clear distinction in the used terminology.

The reason for foodcrusts being an extremely challenging target is that their formation requires the exposure of organic materials to substantial heat. It is not clear what the minimum temperature is at which foodcrust forms. In archaeological foodcrust it is not uncommon to find degradation markers that only form after heating at 270  C for one hour or 200  C for five hours (Bondetti et al. 2021a). Exposure to these levels of temperatures has substantial effects on proteins. Cysteine bridges have been observed to break as a result of

roasting and frying (Hu et al. 2017) and heating also induces a loss of carboxyl groups (Yu et al. 2017). More substantial protein transformations also occur due to the Maillard reactions, leading to a loss of hydrolysable amino acids (Styring et al. 2013).

1.1 The merits of foodcrust

The protein degradation and modification are likely to complicate the extraction of proteins from foodcrust, as well as their analysis. Nonetheless, foodcrusts have the potential to yield valuable information regarding dietary practices in the past. Although the cooking process severely damages proteins, the advantage is that it can be assumed with some confidence that any tissues or taxa detected in the foodcrusts were cooked. The choice to cook certain foods is one of the many ways in which cuisine is expressed and can be identified archaeologically, as it is an example of a cultural practice for processing foods (Graff 2020; Lundy et al. 2023). Especially by contrasting the overview of exploited resources to the selection of foods that were cooked, it becomes possible to identify past culinary practices. The lipid analysis of foodcrust from Jōmon pots from Hokkaidō has identified that most vessels were used for the processing of aquatic resources, while the general economy also included substantial amounts of walnuts and deer (Robson et al. 2020). Another example from the lipid analysis of foodcrusts and ceramic samples from Zamostje 2 is the focus on the cooking of aquatic resources in pottery that appeared in the Middle Neolithic. Older samples from the earlier Neolithic at this site show a larger diversity in processed resources and the faunal assemblage indicates that fish formed a substantial component of the diet from the Mesolithic onwards. It is hypothesised that the dominance of aquatic resources in the lipid residue in the Middle Neolithic does not reflect a change in diet, but rather in how certain foods are prepared (Bondetti et al. 2020). Consequently, rather than only being able to discuss the inclusion of species within the diet, the biomolecular analysis of foodcrust can inform us about the culinary practices of past societies.

Secondly, many of the materials commonly analysed in archaeology to study diet are biased towards certain types of resources. Faunal remains naturally only concern the animal component of the diet, while botanical remains only provide evidence of plant consumption. On the other hand, foodcrust can contain evidence of both the animal and plant components of the diet, allowing for a more comprehensive view of the cooked resources.

Lastly, foodcrusts form an ideal bridge between relative typochronologies and absolute radiocarbon dating (Seegerberg et al. 1991). Ceramic typology plays an important role as a chronological marker in many different archaeological contexts (Trampota and Květina 2020). In Eastern Europe the start of the Neolithic is even defined by the appearance of pottery (Courel et al. 2020). Despite its historically important role in cultural classification, its inorganic composition has made it complicated to correlate cultural chronologies with absolute dates, although recent developments have succeeded in directly dating the organic residues within the ceramics (Casanova et al. 2020). The organic compounds in foodcrusts allow the gap between the two chronological frameworks to be bridged, which is especially valuable for sites where the cultural association relies heavily on ceramic typology.

1.2 Analysing foodcrust

The combination of faunal and plant resources in foodcrusts and its relevance to culinary practices makes foodcrust an interesting material for analysis. Apart from proteomics, the focus of this thesis, foodcrusts are frequently analysed with microscopy (in particular Scanning Electron Microscopy) and lipid analysis. Each of these three methods has a different target of analysis and as a result, provides information regarding a different aspect of the foodcrust.

1.2.1 Scanning electron microscopy (SEM)

SEM has frequently been applied to foodcrusts (Bondetti et al. 2021b; Demirci et al. 2020; Heiss et al. 2020; Kubiak-Martens 2016; Kubiak-Martens et al. 2024; Kubiak-Martens 1999, 2002; Bishop et al. 2022). The main focus of SEM studies on foodcrust has been the identification of plant remains (Hansson and Isaksson 1994; Kubiak-Martens, Brinkkemper and Oudemans 2015; González Carretero, Wollstonecroft and Fuller 2017), and regarding the identification of especially plant underground storage organs the application of SEM has been described as a breakthrough (Kubiak-Martens 2016).

SEM is often able to make species-level identifications of plant remains, although the taxonomic specificity greatly depends on the specific microfossils observed and thus varies significantly between samples (Ge et al. 2024; Kubiak-Martens, Brinkkemper and Oudemans 2015). Beyond taxonomic identification, SEM analysis may be able to determine the manner of cooking based on morphological changes. For example, methods for distinguishing boiling, baking and fermentation of cereals have been proposed (Arranz-Otaegui et al. 2018; González Carretero, Wollstonecroft and Fuller 2017; Heiss et al. 2020; Kubiak-Martens et al. 2024). Apart from plant remains, animal remains such as bone fragments, fish scales or fin rays are sometimes observed during SEM analysis, but these can often only be identified to broad taxonomic groups (Bondetti et al. 2021b; Teetaert et al. 2024).

Compared to lipid analysis and proteomics, SEM has the lowest operational costs. It requires no reagents nor do the foodcrusts need to be chemically altered. In some cases, the surface of the object is coated before analysis, but this is not always necessary. Depending on the size of the target object and the size of the SEM chamber, subsampling may be required, which would make it a form of destructive analysis. One of the defining qualities of SEM is that rather than visible light it uses an electron beam to image an object, which in general allows SEM to obtain higher magnifications and a greater depth of field (Frahm 2014). The electron beam is focused on the target object and the fired electrons can cause electrons from the target to be excited, resulting in the emission of secondary electrons (Zhou et al. 2006). Due to the low energy of secondary ions, usually between 3-5 eV, they are only emitted from nanometers away from the electron beam's target. Consequently, their detection provides highly detailed information on the topography of the surface (Zhou et al. 2006; Frahm 2014). Apart from secondary ions, backscattered electrons also provide information regarding the target object. Backscattered electrons are electrons from the electron beam that hit an atomic nucleus and ricochet back from the target. They are high energy (>50 eV) and the number of backscattered electrons depends on the atomic element the electron hits. The more protons in the nucleus, the more electrons scatter back (Zhou et al. 2006). As a result the contrast in backscattered electron images provides information about differences in the elemental composition of a material (Frahm 2014).

1.2.2 Lipid analysis

In terms of the biomolecular analysis of foodcrust, the most popular target molecules are not proteins, but lipids. Lipid analysis has a long and successful tradition of the analysis of both foodcrusts as well as residues absorbed in the ceramic matrix (Heron and Evershed 1993). It played an important role in the investigation of the use of dairy resources by Neolithic farmers (Copley et al. 2003; Craig et al. 2005).

The most commonly used workflow in recent studies on the analysis of lipids from foodcrust consists of two parts: GC-MS (*Gas Chromatography - Mass Spectrometry*) analysis in order to detect the presence of biomarkers (Evershed 2008) and GC-C-IRMS (*Gas Chromatography - Combustion - Isotope Ratio Mass Spectrometry*) for compound-specific isotope analysis. Biomarkers have been established for a variety of resources at different levels of specificity. Perhaps the most relevant category of biomarkers for the analysis of foodcrust are the aquatic biomarkers, as faunal assemblages are often biased against fish (Zohar et al. 2008; Szpak 2011). Additionally, detecting aquatic resources has been one of the focus points of archaeological organic residue analysis (Craig et al. 2007; Heron and Craig 2015; Demirci et al. 2020; Admiraal et al. 2023a). Aquatic biomarkers include isoprenoid fatty acids and APAAs (ω -(*o*-alkylphenyl) *alkanoic acids*) with C18 and C20 carbon atoms (Evershed et al. 2008; Copley et al. 2004; Bondetti et al. 2021a). Beyond marine resources, there have been recent efforts to characterise biomarkers for cereals, such as alkylresorcinols as indicators of *Triticum* spp. (Colonese et al. 2017; Doliente et al. 2024) and miliacin for broomcorn millet (*Panicum miliaceum*) (Jacob et al. 2008; Heron et al. 2016). Biomarkers can be useful, but require extensive validation as there can be questions regarding their exclusivity, as demonstrated by the search for wine biomarkers (Drieu et al. 2020).

Compound specific isotope analysis via GC-C-IRMS usually focuses on the $\delta^{13}\text{C}$ of palmitic ($\text{C}_{16:0}$) and stearic ($\text{C}_{18:0}$) fatty acids. By plotting the measured $\delta^{13}\text{C}$ ratios of these two fatty acids against each other some broad categories of foodstuff can be distinguished, such as dairy, (non-)ruminant adipose tissue, porcine fats, freshwater resources and marine resources (Cubas et al. 2023). Bulk isotope analysis of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ is often applied in combination with compound-specific isotope analysis on foodcrusts (Craig 2004). The isotopic ratios, both bulk and compound specific, are composites of the isotopic compositions of the various foodstuffs the foodcrust consists of. However, the final measured isotopic ratios of a foodcrust are not direct averages of their mix of foodstuffs. Not all foodstuffs contain an equal amount of lipids and those that contribute more palmitic and stearic fatty acids to the foodcrust will be overrepresented in the compound specific isotope ratios (Heron and Craig 2015). As for the bulk isotope measurements, proteins contain more nitrogen relative to lipids or carbohydrates and thus the $\delta^{15}\text{N}$ will be more reflective of the protein component of the contents (Craig et al. 2007).

A substantial challenge of the interpretation of foodcrust isotopic ratios, both bulk and compound specific, is the potential of the foodcrusts being mixtures of different products. The isotopic evidence may sometimes allow a sample to be identified as a mixture, but it is not possible to directly determine how many resources were combined, in what proportions and what they were. There have been attempts to infer what combination of resources and in which proportions could have resulted in an observed isotopic ratio using mathematical models. A wide variety of mixing models have been applied to archaeological samples, from linear mixing models (Ben-David, Flynn and Schell 1997; Yoneda et al. 2019) to simple probabilistic (Phillips and Gregg 2003; Miller et al. 2020) and Bayesian models (Cheung and

Szpak 2021; Fernandes et al. 2014). A comparison of various mixing models on a theoretical case study showed that they can indeed accurately infer the relative contribution of different foodstuffs to a mixture, but also warned that isotopically similar foodstuffs cannot be distinguished (Cheung and Szpak 2021). An important note to add is that the isotopic composition can vary within a species between different tissues (Codron et al. 2012), which is especially relevant to materials like foodcrust, as they could contain any sort of cooked foodstuffs.

Although these biases complicate the interpretation of isotope ratios, they can be used to our advantage. The larger the proportion of lipids in a mixture the more similar the bulk $\delta^{13}\text{C}$ and compound-specific $\delta^{13}\text{C}$ values should be, while large differences in $\delta^{13}\text{C}$ ratios between bulk and compound specific samples indicate the presence of something other than lipids contributing carbon to the mixture (Robson et al. 2022).

The combination of biomarkers and isotopic analysis has been frequently and successfully applied to a wide range of foodcrusts in recent years and has been especially valuable in providing evidence for the continued use of aquatic resources during the Neolithic (Courel et al. 2020; Papakosta, Oras and Isaksson 2019; Heron and Craig 2015; Lucquin et al. 2016; Shoda et al. 2017) as well as the spread of dairy consumption (Heron et al. 2015; Demirci et al. 2021; Cubas et al. 2020; Robson et al. 2021; Cramp et al. 2014).

Additionally, a recent study pioneered the application of MALDI-MS for the analysis of triacylglycerols (TAGs) as a way to improve the taxonomic resolution of lipid analysis. Initial results seem promising and further research will show how broadly applicable this method is (Drieu et al. 2024).

1.2.3 Palaeoproteomics

Palaeoproteomics, the study of ancient proteins, will be the main focus of this thesis. In theory, proteomics does not share many of the limitations of lipid analysis, the two most important being limited taxonomic resolution and identification of a mixture's components. However, the application of proteomics to ancient dietary mixtures is relatively recent (Buckley, Melton and Montgomery 2013), especially for foodcrust (Colonese et al. 2017; Shevchenko et al. 2018).

Although the earliest studies of proteomics applied to foodcrust are relatively recent, researchers have been trying to recover dietary proteins for a much longer time. The initial attempts focused on ceramic rather than foodcrust (Evershed and Tuross 1996), as it was expected that the mineral matrix of the ceramic would help preserve the proteins (Craig and Collins 2002). There have indeed been successful studies on proteins from ceramic, but their extraction remains very challenging, preservation is often poor (Barker et al. 2018) and the number of recovered proteins is limited (Solazzo et al. 2008; Dallongeville et al. 2011; Chowdhury, Campbell and Buckley 2021; Chowdhury et al. 2022).

1.2.3.1 Protein extraction

Due to the challenges of extracting proteins from ceramic, foodcrust was considered as a more promising alternative. Both materials can provide information regarding diet and cuisine, but foodcrusts are an organic matrix rather than an inorganic matrix. Consequently, it was expected that the extraction of proteins from foodcrust would require less severe chemicals than ceramic (Craig and Collins 2000). There are many different protein extraction protocols published, but in general they all contain the same four steps:

1. Extraction

- a. Reduction and alkylation (optional)
 - b. Purification (optional)
2. Enzymatic digestion
3. Acidification
4. Clean up

In palaeoproteomic studies on bone or other mineralised tissues such as enamel or dental calculus, the extraction step is preceded by a demineralisation step. As foodcrust is not mineralised, demineralisation is not required. The main purpose of the extraction step is to bring whatever proteins are present in the sample into solution. A variety of reagents can be used for this purpose. Commonly used examples include GuHCl (*guanidine hydrochloride*) (Hughes et al. 2019; Cleland 2018; Evans et al. 2024), SDS (*sodium dodecyl sulfate*) (Shevchenko et al. 2018), urea (Lyu et al. 2024) or commercial options such as M-PER (*Mammalian Protein Extraction Reagent*, Thermo Fisher) (Colonese et al. 2017). All these extraction reagents have different chemical properties, but their common characteristics are that they denature and solubilise proteins.

Following incubation in an extraction buffer there are two frequently employed, but optional steps: a) reduction and alkylation and b) purification. Protocols solely focused on the extraction of collagen for peptide mass fingerprinting (i.e. ZooMS) often leave out this step, but most extractions for LC-MS/MS analysis do perform a reduction and alkylation step. The purpose of the reduction step is to break open any disulfide bridges between cysteines. To ensure the disulfide bridges do not reform, a second reagent (the alkylation agent) is added in order to bind an alkyl group to the opened up thiol group on the cysteine residues (Sechi and Chait 1998; Müller and Winter 2017). The most commonly used reduction agents in proteomics are DTT (*dithiothreitol*), TCEP (*tris-(2-carboxyethyl)-phosphine*) and BME (*β -mercaptoethanol*), while common alkylation reagents are IAA (*iodoacetamide*), IAC (*iodoacetic acid*), AA (*acrylamide*) and CAA (*chloroacetamide*) (Müller and Winter 2017). Although the above reagents have all been commonly used, a recent review showed that IAA and IAC alkylation resulted in poor protein identification rates compared to the other alkylation agents (Müller and Winter 2017).

If a reduction and alkylation has been carried out it is necessary to perform some form of purification as the reduction and alkylation agents can disrupt the enzymatic digestion. Additionally, purification might be performed in order to increase the protein concentration and to remove any non-proteinaceous material from the solution. The presence of non-proteinaceous material can have negative effects on protein identification, either by adding noise to the mass spectra or by clogging up the chromatographic column. Commonly used purification methods that have been applied in palaeoproteomic studies can be subdivided into the following four categories:

1. Filter-aided separation protocols (FASP) (Wiśniewski et al. 2009)
2. Gel-aided separation protocols (GASP) (Fischer and Kessler 2015)
3. Covalent binding separation, (e.g. SP3) (Hughes et al. 2014)
4. Chromatographic separation (e.g. C18 columns) (Rappsilber, Mann and Ishihama 2007)

FASP employs molecular weight cut-off filters to retain high molecular weight compounds, such as proteins, while low molecular weight material flows through. After the filtration enzymatic digestion is performed on the filter, which cleaves the proteins into small enough peptides to flow through the filter and be retrieved (Wiśniewski 2019).

GASP relies on strong detergents such as SDS to denature the proteins, which are then stabilised in an acrylamide gel via copolymerisation. Gels have been found to be quite

effective at protein purification, but require extensive clean up later, as small amounts of detergent can have a large impact on enzymatic digestion, disrupt the chromatographic column and lead reduced peptide identifications (Zhang and Li 2004; Lu and Zhu 2005).

Protocols such as SP3 (Single-Pot Solid-Phase-enhanced Sample Preparation) use coated beads with a magnetic core to filter out non-proteinaceous material. The coating (various types of coating are commercially available) facilitates the binding of proteins to the beads, which are then fixed in place in a magnetic field. Only non-proteinaceous compounds then remain in solution which can be easily taken off and discarded, while the bead-bound proteins are immobilised (Hughes et al. 2019). An advantage of this technique is that it does not require detergents nor additional rigorous cleaning (Hughes et al. 2014). The magnetic properties of the beads facilitate protein purification as well as separation of peptides from the beads after digestion. Additionally, SP3 has been the focus of a number of studies explicitly optimising the protocol for ancient (bone) proteins (Cleland 2018).

Another way to concentrate the proteins in an extraction is via chromatography (Gobom et al. 1999). Pipette tips with a resin bead packed filter, such as ZipTips or StageTips (Rappsilber, Mann and Ishihama 2007), are frequently used for the clean up of samples, but can also be used in an initial concentration step. The exact principle of purification depends on the filter used and multiple types of filter can be combined for combined purification. The most commonly used filter in palaeoproteomics is C18 resin, which is a hydrophobic compound, enabling the facile washing of all non-proteinaceous material. Later, the proteins can easily be released from the filter by eluting with an organic solvent (Rappsilber, Mann and Ishihama 2007).

Following purification or immediately after extraction the proteins are typically enzymatically digested. Proteins are digested in order to 'standardise' peptides, in the sense that by cutting consistently after the same amino acids a limited and predictable set of peptides is generated, which facilitates downstream analysis. There are (modern) proteomic studies that do not digest the extracted proteins. These studies are called top-down as they focus on the complete proteins (Melby et al. 2021). In palaeoproteomics no complete proteins are expected to be recovered as a result of fragmentation due to diagenesis. Consequently, palaeoproteomic studies are bottom-up and typically include a digestion step to standardise peptide formation (Warinner, Korzow Richter and Collins 2022). There are some exceptions to the use of proteases (Nielsen-Marsh et al. 2002), the primary being enamel protein extraction (Cappellini et al. 2019), although there are examples of enamel protein analysis including digestion as well (Parker et al. 2019). The reason for enamel extractions usually not including a digestion step is that enamel proteins are already digested *in vivo* and thus do not need to be digested later (Bartlett 2013). As for all other sample types that are usually subjected to enzymatic digestion, trypsin is the most commonly used protease. Trypsin cleaves protein sequences after each lysine (K) and arginine (R) (Rodriguez et al. 2008). Traditionally, trypsin digestions have been left overnight, but recent research shows that digestion duration as short as three hours may perform equivalently (Le Meillour et al. 2024).

Recent research on archaeological bone has also started to explore the use of alternative proteases, either independently or in combination with trypsin (Wilkin et al. 2024). These studies observed that of the tested proteases, Glu-C and chymotrypsin led to the largest increase in number of identified proteins and protein coverage, both when used in parallel to trypsin (Lanigan et al. 2020) or consecutively (Fagernäs et al. 2024).

After digestion, the samples are usually acidified in order to halt the enzymatic activity. Especially in the case of trypsin, reduced cleavage specificity has been observed to

occur if left unchecked (Keil-Dlouha et al. 1971; Vandermarliere, Mueller and Martens 2013). A final-clean up step, usually either using ZipTips or StageTips is also performed to remove any salts from the solution. Palaeoproteomic protocols often feature multiple cleaning steps, because there is much less control over the purity of the sample compared to modern cell digests for example. However, no purification step has perfect efficiency and there is always the risk of protein loss. Therefore determining the rigour of cleaning is a tradeoff between protein loss and increased protein recovery due to the removal of noise-inducing and column-clogging material.

1.2.3.2 Mass spectrometry data acquisition

Once the digested peptides have been acidified and cleaned from all non-proteinaceous material they can be analysed with a mass spectrometer. There are many ways to categorise mass spectrometry techniques and their properties but for the purpose of palaeoproteomics the most important distinction is between single mass spectrometry (MS) and tandem mass spectrometry (MS/MS). In MS analysis the peptides are ionised and their mass to charge ratio (m/z) is measured. Based on the m/z values of peaks observed in the mass spectrum the identity of the dominant peptides can be established. This approach is called PMF (*Peptide Mass Fingerprinting*), as the group of peaks diagnostic of a protein resembles a fingerprint or barcode (Ostrom et al. 2006; Collins, Buckley and Grundy 2010). In palaeoproteomics, ZooMS (*Zooarchaeology by Mass Spectrometry*) (Buckley et al. 2009), is the most common application of PMF and of palaeoproteomics in general as well. ZooMS focuses on taxonomic identification using sequence differences in collagen type 1. ZooMS data acquisition is usually done with a *Matrix Assisted Laser Desorption/Ionisation - Time Of Flight - Mass Spectrometer* (MALDI-TOF-MS), but neither a MALDI source nor TOF detector are strictly speaking required. The primary reason for the MALDI-TOF-MS's popularity is its speed and resulting low analytical cost (Thiede et al. 2005), which makes it a perfect high-throughput tool for the taxonomic screening of large batches of samples.

The other main application of palaeoproteomics is shotgun proteomics, which relies on various configurations of *Liquid Chromatography - tandem Mass Spectrometry* (LC-MS/MS). The key difference between PMF and shotgun proteomics lies not in the use of MALDI or LC nor in the choice of detector, rather in the addition of a second round of fragmentation (Thiede et al. 2005). After the initial measurement of the m/z a number of ions are selected for fragmentation, which then occurs and the m/z values of the fragments are measured. The coupling of the full peptide (the precursor) m/z with the fragment m/z values provides greater confidence in identifying the sequence and associated protein and taxon of the observed ion (Marcotte 2007). In LC-MS/MS studies this process is complemented by the gradual release of precursors from the LC column. A sample is injected onto the column and the column is washed by a solution with an increasing gradient of polarity, resulting in a gradual release of peptides separated by their polarity. The peptides are then ionised, frequently via *Electrospray Ionisation* (ESI) and sent to the first mass detector. After the acquisition of the precursor mass spectra the selection of precursors to be fragmented can be made in different ways. Broadly speaking, in *Data Dependent Acquisition* (DDA), a number of the most intense ions are selected for fragmentation, while in *Data Independent Acquisition* (DIA) all ions within a particular m/z window are selected for fragmentation (Doerr 2015). DIA unbiased selection for fragmentation results in improved reproducibility in terms of which ions are selected for fragmentation (Bilbao et al. 2015), but it can happen that

co-eluting peptides are fragmented together, causing the MS/MS spectra to derive from multiple precursors (Pino et al. 2020). At the moment, DDA is by far the most frequently used approach in palaeoproteomics (Warinner, Korzow Richter and Collins 2022), although DIA has been applied to archaeological bone (Rüther et al. 2022; Hansen et al. 2024; Ma et al. 2024).

The selected precursors can then be fragmented in several ways, which will determine the type of fragment ions that form. Some of the most commonly used fragmentation methods are:

- *Collision-Induced Dissociation (CID)*
Precursor ions are made to collide with an inert gas and the resulting energy from the collisions break the weakest bonds. If the broken bond is the backbone amide bond, this leads to the formation of b- (N-terminal) and y-ions (C-terminal) (Sobott et al. 2009).
- *Higher-energy Collision Dissociation (HCD)*
HCD functions similar to CID, except the collisions occur under a higher voltage (Olsen et al. 2007). It also creates b- and y-ions.
- *Electron Capture Dissociation (ECD)*
In ECD the energy required to fragment the precursor ion comes from the capture of free electrons, which subsequently lead to the breakage of the bond between the nitrogen of the peptide bond and the first following carbon atom. The resulting fragment ions are c- (N-terminal) and z-ions (C-terminal) (Qi and Volmer 2017).
- *Electron Transfer Dissociation (ETD)*
Similar to the difference between CID and HCD, the mechanism of ETD is very similar to that of ECD. The main difference is that in ETD the electrons are provided by anions. It also results in the formation of c- and z-ions (Mikesh et al. 2006).

The core difference between CID and HCD on the one hand and ECD and ETD on the other, is that the collision-based approaches, such as CID and HCD, not only fragment the peptide backbone, but can also lead to the loss of *Post Translational Modifications* (PTMs) (Qi and Volmer 2017).

Currently, the two commonly used frontrunners in mass spectrometry are Orbitrap-detector mass spectrometers (Thermo Scientific) and the timsTOF (Bruker). The defining feature of Orbitrap-style mass spectrometers is the Orbitrap, which is an analyser and a detector. The Orbitrap consists of two outer and one central electrode, which cause ions to oscillate between the two electrodes on the outer ends of the analyser. The frequency of an ion's oscillations between the electrodes is dependent on the ion's m/z (Hu et al. 2005). The Orbitrap measures an ion's frequency and then applies Fourier transformation in order to convert the detected frequency to a m/z value (Hu et al. 2005; Eliuk and Makarov 2015). The end result of this complex process is a mass spectrometer with an extremely high resolution and mass accuracy (Martins et al. 2016).

While the defining feature of Orbitrap is the analyser and detector, the timsTOF stands out due to its additional ion separation using ion mobility. After ionisation, the ions enter a tunnel where they are pushed forward by a gas while being held in place by an electrical field. The strength of the electrical field is then gradually decreased, causing larger ions with a greater cross section to progress in the mass spectrometer before ions with a smaller cross section (Michelmann et al. 2015). The primary purpose of the ion mobility in a timsTOF is to improve the resolving power of the mass spectrometer (Michelmann et al.

2015). A secondary side effect is that timsTOF mass spectrometers are much faster than Orbitrap-style mass spectrometers, which is mainly due to the difference in MS/MS acquisition rates between Orbitrap and TOF mass analysers (Meier et al. 2018; Hao et al. 2012).

1.2.3.3 Bioinformatic data analysis

The bioinformatic analysis of proteomic data acquired as part of ZooMS is relatively straightforward. The spectra can be visualised using software such as Mmass (Strohalm et al. 2010) and taxonomic identification is based on the comparison of manually identified peaks with lists of known biomarkers, such as a collection of published biomarkers hosted by the University of York (Presslee 2020). There have been attempts to (semi-)automate the process of biomarker-based taxonomic identification (Gu and Buckley 2018; Hickinbotham et al. 2020; Végh and Douka 2024), but for now manual identification remains the standard.

Analysing data from shotgun proteomics experiments is a more complex process and involves several choices that have a substantial impact on the resulting protein and taxonomic identifications. A simplified description of the protein identification process is that the observed precursor and fragment peaks are compared to a database of *in silico* digested proteins. Potential matches between observed peaks and peptide sequences are scored based on how well they fit and, among others, the presence of PTMs, missed or semi-specific cleavage and the extent of fragment ion coverage (Cottrell 2011). The top scoring peptide (and associated protein and species) are then assigned as the *Peptide Spectrum Match* (PSM) (Granhölm and Käll 2011).

A first important choice in the protein identification process is which protein identification programme to use. There are many different options available, including proprietary and freely available software. There is no single standard programme, but software like Mascot (Perkins et al. 1999) and MaxQuant (Cox and Mann 2008) are among the most commonly used programmes. Some alternative programmes with distinctive features include Casanovo (Yilmaz et al. 2024), PEAKS (Zhang et al. 2012) and Novor (Ma 2015) for *de novo* sequencing and MetaMorpheus (Wenger and Coon 2013; Solntsev et al. 2018), pFind3 (Chi et al. 2018) and PeptideProphet (Keller et al. 2002) for *Post Translational Modification* (PTM) discovery. PeptideProphet can also be used as part of the Fragpipe pipeline, which facilitates the combination of several mass spectrometric software programmes for protein identification, validation and quantification (Kong et al. 2017; Yu et al. 2020; Yu, Deng and Nesvizhskii 2024; Yu, Haynes and Nesvizhskii 2021; da Veiga Leprevost et al. 2020). A systematic review of all the available software is beyond the scope of this introduction, especially as it would require empirical testing. Due to the uniqueness of the scoring algorithms used by each programme differences in performance are expected even between programmes with similar features. The magnitude of these differences is difficult to predict. Which programme will be most suited will depend on what features are most important considering the dataset, such as robust quantification for expression analysis and PTM discovery for samples that likely contain many PTMs.

The second impactful choice to be made is the selection of a protein sequence database. Only proteins that are featured within the protein database can be identified, so it is important to ensure all potential proteins present in the sample are also represented in the database. It is vital that potential contaminants are included in the database, as otherwise the presence of contamination may go unnoticed (Hendy et al. 2018a). However, if the database is too large in comparison to the number of proteins present in the samples, PSM

identification may be negatively affected (Jeong, Kim and Bandeira 2012; Rodriguez Palomo et al. 2024).

Apart from selecting a protein search engine and database an important decision is determining the criteria used for accepting protein identifications. Some of these settings need to be set in the protein search engine, such as *False Discovery Rate* (FDR), others are used to filter the output protein identifications. It is standard practice to exclude any proteins represented only by a single peptide (Hendy et al. 2018a), but there is some variation in more strict acceptance criteria. For example, some studies require two unique peptides for accepting a protein identification (Welker et al. 2020) and others three (Chowdhury, Campbell and Buckley 2021). The two-peptide requirement is as arbitrary as the commonly used p-value threshold of 0.05 and has received criticism (Gupta and Pevzner 2009). However, there is no clear process for establishing the minimum number of peptides needed to confidently identify a protein in palaeoproteomics. One of the primary obstacles to establishing a minimum threshold in palaeoproteomics is the challenge to find a protein that is known to be present in the sample and whose abundance can be used as a minimum. Even in the case of proteins that are expected in a sample, based on our understanding of the proteomic composition of the material during life, their detection does not necessitate their authenticity, as the original protein may have degraded and the detected PSMs may be false positives. Considering the lack of a data-drive minimum peptide threshold, the second best option is to facilitate interstudy comparison by upholding conventional identification standards, such as the two-peptide rule.

Many palaeoproteomic studies include a last step to validate the taxonomic identifications assigned by the protein identification software. The reason for not initially accepting the protein search engine's taxonomic identification is twofold: first, the database used in the protein search is often incomplete and a validation against a wider database may help in refining the taxonomic identification. Secondly, a validation step ensures that the associated peptides are not shared between multiple species. The most popular tool for taxonomic validation is BLAST (Altschul et al. 1990), but recently additional tools such as Unipept (Gurdeep Singh et al. 2019) and ProteoClade (Mooradian et al. 2020) have started to be used more as well.

Even if a protein fulfils all identification criteria and its taxonomic identification has been validated it does not necessarily mean that the protein is ancient. There is no strict threshold for establishing the authenticity of ancient proteins but the percentage of deamidated peptides of a protein has been used as a way to distinguish ancient from modern proteins (Wilson, van Doorn and Collins 2012; Robinson and Robinson 2001). Deamidation is a PTM which sees glutamine transform to glutamic acid and asparagine to aspartic acid. Previous studies have established that modern samples are in general less deamidated than ancient samples, but there can be overlap and there is no clear cutoff value that guarantees a protein to be ancient (Schroeter and Cleland 2016; Brown et al. 2021; Ramsøe et al. 2020). Nonetheless, the absence of deamidation or substantially lower values in a selected number of proteins compared to the sample as a whole, may be a useful indicator of questionable authenticity (Ramsøe et al. 2020).

1.3 The archaeological relevance of foodcrust

Due to the ubiquity of ceramics in the archaeological record, the biomolecular analysis of organic residues can be applied to the majority of Holocene archaeological contexts and even some Pleistocene (Craig et al. 2013; Lucquin et al. 2016). As such,

organic residue analysis is able to inform on questions related to diet and cuisine for the majority of Holocene archaeological cultures. Notwithstanding the range of its applications, it can be argued that the main contributions of biomolecular organic residue analysis have focused on three debates:

1. The use of the earliest pottery and consequently the reasons for adopting pottery.
2. Culinary specialisation.
3. Continuity and change in diet and cuisine during periods of cultural changes.

Why ceramics emerged and became adopted near-universally has been a longstanding archaeological debate. One perspective on this debate focuses on the use of the first ceramic objects. Large and wide scale application of organic residue analysis may reveal common tendencies in the use of ceramics by populations adopting the technology. Such cross-cultural patterns may be indicative of common reasons for adopting pottery, which might inform us on why ceramic technology became so widespread. An example of such a cross-cultural pattern is the identification of aquatic resources in the earliest pottery found in Korea (Shoda et al. 2017), Japan (Craig et al. 2013; Lucquin et al. 2016; Robson et al. 2020) and Alaska (Admiraal et al. 2023b) using lipid analysis. This preference for aquatic resources seems to occur across a large geographical area and is also found in the Eastern Baltic, among the ceramic-using hunter-gatherers of the Narva culture (Papakosta, Oras and Isaksson 2019; Oras et al. 2017; Courel et al. 2020). However, lipid analysis, which has been used to find these patterns, has its own limitations (section 1.2.2). The most obvious perhaps being its bias towards lipid-rich foods. Consequently, patterns in usage are likely to gravitate towards the use of lipid-rich resources, as they are easiest to identify. Therefore, it is important to also employ other methods to study the organic residues, such as SEM (section 1.2.1) and proteomics (section 1.2.3). A holistic approach to diet is necessary when searching for patterns in use, as the biases of each method may otherwise influence the resources that can be identified.

The same applies, perhaps even more so, to the study of specialisation in food processing. The consistent selection of a small portion of the available foodstuffs for particular types of food processing, such as cooking, is another major archaeological debate that biomolecular organic residue analysis has been applied to. For example, lipid analysis suggested that Swifterbant ceramics were used primarily for the processing of freshwater fish (Demirci et al. 2020, 2021) and proteomic analysis has been used to confirm that a type of pot similar to modern cheese strainers, is likely to have been used as a cheese strainer in the past as well (Salque et al. 2013)(Hendy et al. 2018b; Fonseca, Freeman and Collins 2022, 2024). Considering the different biases each of the commonly used methods, such as lipid analysis, SEM and palaeoproteomics, have, caution must be exercised in claiming evidence for specialisation. Only through a combination of methods can a more complete view of the food processed in pottery, and thus potential specialisation, be obtained.

Although popular research themes, the study of the first use of pottery and specialisation of pottery use, encounter a number of challenges when studied using organic residue analysis as described above. Palaeoproteomics can contribute to these topics, but faces similar issues. The third debate, the study of economic transitions, might fit better with the advantages of palaeoproteomics. Changes in diet occurring alongside cultural changes is a frequent topic of research in archaeology and it is not reserved for organic residue analysis, or proteomics. However, the additional resolution of proteomics and its ability to distinguish components of a mixture can help to uncover if there are differences in which specific foods are cooked in ceramic vessels, even if the larger picture of the faunal and botanical assemblage suggests that the overall diet remains the same.

Some cultural transitions are especially suited for organic residue analysis as they feature the introduction of new foods. The speed and scale of the adoption of these new resources in the existing diet can be easily studied with organic residue analysis (Lundy et al. 2023). The Neolithic-Mesolithic transition in Western Europe is an especially suitable period, as it is marked by the introduction of a wide range of foods, including cereals and dairy, that were not available to the preceding hunter-gatherers (Sørensen and Karg 2014; Gron, Montgomery and Rowley-Conwy 2015). Furthermore, some of the most typical foods for either side of the transition, aquatic resources for the Mesolithic and dairy for the Neolithic, should be easily detectable using organic residue analysis. Dairy proteins, as mentioned before, are frequently observed with lipid analysis and proteomics, while aquatic resources are commonly identified using lipid analysis. Additionally, the protein richness of fish muscle may be beneficial to the chances of recovering fish proteins, but little is known regarding the potential of fish muscle proteins to withstand the degradation induced by heating and burial.

1.4 Objectives

The combination of animal and plant evidence with direct evidence of cooking makes foodcrust an attractive material for studying diet, but it is unclear how the matrix of foodcrust and its exposure to cooking and burial impacts the reliability and precision of proteomic taxonomic identification. The main aim of this thesis is to clarify and address this uncertainty regarding the potential and limitations of the proteomic analysis of foodcrust. In particular, two questions are central to the thesis:

- To what extent do the resources identified through proteomic analysis fairly represent the foods cooked to produce the foodcrust?
- How accurate and precise are the taxonomic identifications of proteins from foodcrust?

These two core questions form a red thread throughout the thesis. They are the underpinning questions that guided the experiment design of chapters 3 and 4 and the answer to these questions determines how the results of the analyses of the archaeological foodcrust presented in chapter 5 should be interpreted.

Although the main focus of this thesis is the study of foodcrust, chapter 2 stands out from the other chapters in this thesis by focusing on a different material: bone. Despite dealing with bone rather than foodcrust, chapter 2 still plays an important role in the structure of the thesis. Its two main aims are:

- To provide a baseline for the taxonomic resolution achievable with palaeoproteomic analysis in beneficial conditions (i.e. single-species object, relatively good protein preservation).
- To compare the taxonomic resolution of multiple palaeoproteomic workflows.

By comparing the taxonomic resolution of the protein analysis provided by multiple workflows, it becomes clearer how to maximise the precision of the taxonomic identifications. Using the results from bone as a baseline for the analysis of the foodcrust samples helps to

distinguish the limitations of the material from the limitations of the method.

The other chapters too have, apart from the core questions, their own, more specific aims. Chapter 3, an experiment of forming foodcrust using mixtures of carbohydrate-rich, protein-rich and lipid-rich foodstuffs, aims to start addressing some of the many still present fundamental questions regarding the nature and formation of foodcrust. In particular, chapter 3 intends to:

- Determine if there are biases in the recovery of proteins from lipid-rich, protein-rich or carbohydrate-rich resources compared to each other and if these differences persist through multiple cooking events and burial.
- Test if all ingredient combinations result in an equal volume of foodcrust forming, or what conditions result in the largest volume of foodcrust.

With these aims chapter 3 intends to address the formation of foodcrust and how it shapes the evidence that we may be able to recover from foodcrust. It may be that the abundance of foodcrust, or its absence, could already provide some information regarding the contents of the vessel.

Chapter 4 is a test of data analysis and to a lesser extent protein preservation. The chapter discusses the analysis of two sets of experimentally produced foodcrust, which are analysed as archaeological samples would be. The key difference is that the true contents of these foodcrusts are known and revealed after the analysis. The main aim of the experiment can be described as:

- Testing the accuracy of the protein identifications, including whether a pattern can be observed in which foodstuffs are frequently not identified. However, most importantly if false positives are identified.
- Establishing the taxonomic resolution practically feasible in the analysis of foodcrust.

Together, these two aims form the foundation of the framework to interpreting archaeological foodcrust. The theoretical advantages and potential limitations of foodcrust have been described in section 1.1, but chapter 4 will clarify the extent to which they reflect reality.

Building on the lessons provided by chapters 3 and 4, chapter 5 seeks to test what can be learned from the application of palaeoproteomic analysis to foodcrust. Chapter 5 combines both a methodological and archaeological component. The methodological component aims to highlight how proteomic analysis performs relative to other commonly used methods to study organic residues, while the archaeological component seeks to apply palaeoproteomics to a major cultural-economic change: the Mesolithic-Neolithic transition. The aims can be explicitly summarised as:

Methodological

- Determining the frequency of successful protein recovery from archaeological foodcrust.
- Identifying the presence of systematic differences in the resources that can be identified by SEM, lipid analysis and proteomic analysis.

Archaeological

- To study continuity and change in the cooked resources in the foodcrust between the Mesolithic Ertebølle and Neolithic Funnelbeaker at the sites of MLF00906-I and MLF00906-II.

The results to all the above objectives will culminate in an improved understanding of the palaeoproteomic analysis of foodcrusts, by providing empirical data on the accuracy and precision of the taxonomic identifications and by identifying the blind spots of the method. Furthermore, the combined application with methods like SEM and lipid analysis will demonstrate how combined approaches can lead to a more holistic understanding of the foods cooked in ceramic vessels in the past. Lastly, these insights will enable us to gain a better understanding of the impacts of the Mesolithic-Neolithic transition on food cooking in Lolland, Denmark.

1.5 Thesis outline

This thesis aims to investigate the capabilities and limitations of palaeoproteomics in identifying the taxa that make up foodcrust. By addressing the research questions listed above, this thesis hopes to contribute to the understanding of foodcrusts and their use for the investigation of culinary practices.

Chapter 2:

Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands

There is reason to believe that the taxonomic identification of foodcrust will prove more complicated than is usually the case for the analysis of single-species objects. To establish a baseline of the capabilities and limitations of palaeoproteomic taxonomic identification, a variety of palaeoproteomic techniques is applied in a semi-controlled environment; i.e. a bone tool where there is only a single correct taxonomic identification. The taxonomic accuracy in identifying the taxon of a bone should be considered as an upper limit for the potential of foodcrusts for dietary analysis. Additionally, the analysis of a bone should already highlight a number of challenges that might be encountered during the analysis of foodcrust.

Chapter 3:

An unclean slate, discrepancies between food input and recovered lipid and protein signal from experimental foodcrusts.

One of the reasons that materials such as bone serve as a controlled environment for proteomic analysis is because the formation and biochemical composition of bone is well studied. In contrast, it is not clear what defines foodcrusts nor has it been established what is required for foodcrusts to form. Yet an understanding of the transformation of foodstuffs into foodcrusts and how this impacts the protein content is essential for correctly interpreting any protein identifications obtained from foodcrust. This chapter aims to elucidate the food-foodcrust transformation process by mixing carbohydrate-rich (beetroot), lipid-rich (beef fat) and protein-rich (salmon) foods in varying relative proportions and cooking these

repeatedly in ceramic vessels. This experiment enables the observation of how protein recovery changes with increased cooking and burial. Of equal importance is the relationship between the relative composition of the input foods and the output of the proteins in the foodcrust. Lastly, the absolute volume of the foodcrusts is measured at the end of the cooking experiment in order to determine whether particular combinations of foodstuffs result in a larger volume of foodcrust.

Chapter 4:

Trial by fire, a blind test of the proteomic analysis of experimental foodcrust

A well-known disadvantage of palaeoproteomics is its reliance on incomplete databases, limiting its ability to correctly identify the taxon of a sample. The scope of this issue is relatively well-described for bone, especially for collagen, but this does not translate directly to foodcrusts, considering the variety of tissues and taxa that the proteins recovered from foodcrust can belong to. Additionally, the heating-induced damage and modifications of the surviving protein are expected to complicate protein identification, either by inducing unspecific cleavage, reducing the amount of protein or adding unexpected mass shifts. These issues complicate taxonomic identification and may prevent certain species from being reliably identified, despite their presence in the foodcrust.

Chapter 4 aims to establish the range of taxa that one can expect to accurately identify in foodcrusts by performing a blind test. In this blind test 20 experimentally produced foodcrusts composed of a variety of resources were analysed using two protein extraction protocols, two mass spectrometers and two taxonomic classification tools in order to highlight which taxa remained unidentified.

Chapter 5:

Palaeoproteomic, lipid and SEM analysis of foodcrusts from Syltholm II (MLF00906-I/II), Denmark, reveals dairy consumption in the Mesolithic Ertebølle

Taking into account the limitations for foodcrust analyses as set by the previous chapters, Chapter 5 aims to demonstrate the knowledge that palaeoproteomic analysis can provide on ancient diet by analysing an assemblage of foodcrusts spanning the Mesolithic to Neolithic transition from the Danish site Syltholm II (MLF00906-I/II).

Apart from palaeoproteomics, SEM and lipid analysis were applied to the same assemblage. SEM and lipids are common methods for the study of foodcrusts and are biased towards the plant and fatty component of the mixture respectively. Combining palaeoproteomics, SEM and lipids demonstrates the different resources that each of these methods can identify and which resources they miss. Additionally, the combined approach highlights how these methods complement each other and enables a more comprehensive understanding of what foodcrusts can tell regarding past dietary practices.

Chapter 6: Discussion

Across the previous four chapters several overarching challenges have been identified. Firstly, a review of the literature revealed a lack of standardisation in the

terminology used to discuss foodcrusts, which complicates the comparison of similar materials and the methods used to analyse them. Secondly, the lack of knowledge regarding the biochemical composition of foodcrusts severely impacts our ability to effectively extract proteins from foodcrust.

The multi-species composition of foodcrusts also adds a number of challenges to the bioinformatic analysis, an issue highlighted already in this introduction. The choice of protein reference database and how to construct it has significant effects on the amount of identified peptides as well as on which species can be observed. Moreover, the recent focus on maximising the number of identified peptide-spectrum-matches in protein search engine development may have led to an unintended side effect of spectra being shoehorned into incorrect taxonomic identifications.

Lastly, the discussion focuses on the place of proteomics in dietary analysis as a whole. In particular, how does palaeoproteomics deliver on its promises of tissue-specificity and high taxonomic resolution?

Chapter 7: Conclusion

In the last chapter the insights gained from the different experiments will be combined. The analysis of the bone tool from Bronze Age Heiloo in Chapter 2 showed that palaeoproteomic taxonomic identification can reach species-level specificity, even in cases where no appropriate reference sequences yet exist. Chapter 3 made clear that there is a steep decline in protein preservation after burial and that in most cases it is likely that not all taxa cooked in a vessel can be recovered. The blind test in Chapter 4 showed that the avoidance of false positives is a significant challenge in proteomic analysis. Lastly, the archaeological case study of Syltholm II (MLF00906-I/II) showed that palaeoproteomics could significantly alter our understanding of past diets, but that alternative methods such as SEM and lipid analysis shed light on different aspects of the diet. By itself, palaeoproteomics is insufficient to understand past dietary practices, but the combination of multiple methods is vital for a representative view.

References

- Admiraal, M. et al. (2023a). Chemical analysis of pottery reveals the transition from a maritime to a plant-based economy in pre-colonial coastal Brazil. *Scientific reports*, 13 (1), p.16771. [Online]. Available at: doi:10.1038/s41598-023-42662-5.
- Admiraal, M. et al. (2023b). The role of salmon fishing in the adoption of pottery technology in subarctic Alaska. *Journal of archaeological science*, Elsevier. [Online]. Available at: doi:10.1016/j.jas.2023.105824.
- Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410. [Online]. Available at: doi:10.1016/S0022-2836(05)80360-2.
- Arranz-Otaegui, A. et al. (2018). Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *Proceedings of the National Academy of Sciences*, 115 (31), National Acad Sciences., pp.7925–7930. [Online]. Available at: doi:10.1073/pnas.1801071115.
- Barker, A. et al. (2012). An Optimized Approach for Protein Residue Extraction and Identification from Ceramics After Cooking. *Journal of Archaeological Method and Theory*, 19 (3), pp.407–439. [Online]. Available at: doi:10.1007/s10816-011-9120-5.
- Barker, A. et al. (2018). Taphonomy and negative results: An integrated approach to ceramic-bound protein residue analysis. *Journal of archaeological science*, 94, pp.32–43. [Online]. Available at: doi:10.1016/j.jas.2018.03.004.
- Bartlett, J. D. (2013). Dental enamel development: proteinases and their enamel matrix substrates. *ISRN dentistry*, 2013 (1), Hindawi Limited., p.684607. [Online]. Available at: doi:10.1155/2013/684607 [Accessed 21 October 2024].
- Ben-David, M., Flynn, R. W. and Schell, D. M. (1997). Annual and seasonal changes in diets of martens: evidence from stable isotope analysis. *Oecologia*, 111 (2), Springer Science and Business Media LLC., pp.280–291. [Online]. Available at: doi:10.1007/s004420050236 [Accessed 20 October 2024].
- Bilbao, A. et al. (2015). Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics*, 15 (5-6), Wiley., pp.964–980. [Online]. Available at: doi:10.1002/pmic.201400323.
- Bishop, R. R. et al. (2022). Getting to the root of the problem: new evidence for the use of plant root foods in Mesolithic hunter-gatherer subsistence in Europe. *Vegetation history and archaeobotany*. [Online]. Available at: doi:10.1007/s00334-022-00882-1.
- Bondetti, M. et al. (2020). Fruits, fish and the introduction of pottery in the Eastern European plain: Lipid residue analysis of ceramic vessels from Zamostje 2. *Quaternary international: the journal of the International Union for Quaternary Research*, 541, pp.104–114. [Online]. Available at: doi:10.1016/j.quaint.2019.05.008.
- Bondetti, M. et al. (2021a). Investigating the formation and diagnostic value of ω -(o-alkylphenyl)alkanoic acids in ancient pottery. *Archaeometry*, 63 (3), Wiley., pp.594–608. [Online]. Available at: doi:10.1111/arcm.12631.
- Bondetti, M. et al. (2021b). Neolithic farmers or Neolithic foragers? Organic residue analysis of early pottery from Rakushechny Yar on the Lower Don (Russia). *Archaeological and anthropological sciences*, 13 (8), p.141. [Online]. Available at: doi:10.1007/s12520-021-01412-2.

- Bradfield, J. et al. (2019). Identifying the animal species used to manufacture bone arrowheads in South Africa. *Archaeological and anthropological sciences*, 11 (6), Springer., pp.2419–2434. [Online]. Available at: doi:10.1007/s12520-018-0688-5.
- Brown, S. et al. (2021). Examining collagen preservation through glutamine deamidation at Denisova Cave. *Journal of archaeological science*, 133, Elsevier., p.105454. [Online]. Available at: doi:10.1016/j.jas.2021.105454.
- Buckley, M. et al. (2009). Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 23 (23), Wiley., pp.3843–3854. [Online]. Available at: doi:10.1002/rcm.4316.
- Buckley, M. and Kansa, S. W. (2011). Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. *Archaeological and anthropological sciences*, 3 (3), Springer Science and Business Media LLC., pp.271–280. [Online]. Available at: doi:10.1007/s12520-011-0066-z [Accessed 16 October 2024].
- Buckley, M., Melton, N. D. and Montgomery, J. (2013). Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. *Rapid communications in mass spectrometry: RCM*, 27 (4), Wiley., pp.531–538. [Online]. Available at: doi:10.1002/rcm.6481.
- Cappellini, E. et al. (2019). Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature*, 574 (7776), Springer Science and Business Media LLC., pp.103–107. [Online]. Available at: doi:10.1038/s41586-019-1555-y [Accessed 21 October 2024].
- Casanova, E. et al. (2020). Accurate compound-specific ¹⁴C dating of archaeological pottery vessels. *Nature*, 580 (7804), nature.com., pp.506–510. [Online]. Available at: <https://www.nature.com/articles/s41586-020-2178-z%3C/i%3E>.
- Chen, F. et al. (2019). A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature*, 569 (7756), nature.com., pp.409–412. [Online]. Available at: doi:10.1038/s41586-019-1139-x.
- Cheung, C. and Szpak, P. (2021). Interpreting past human diets using stable isotope mixing models. *Journal of archaeological method and theory*, 28 (4), Springer Science and Business Media LLC., pp.1106–1142. [Online]. Available at: doi:10.1007/s10816-020-09492-5.
- Chi, H. et al. (2018). Open-pFind enables precise, comprehensive and rapid peptide identification in shotgun proteomics. *bioRxiv*, p.285395. [Online]. Available at: doi:10.1101/285395 [Accessed 1 March 2024].
- Chowdhury, M. et al. (2022). Novel Deep Eutectic Solvent-Based Protein Extraction Method for Pottery Residues and Archeological Implications. *Journal of proteome research*, 21 (11), American Chemical Society., pp.2619–2634. [Online]. Available at: doi:10.1021/acs.jproteome.2c00340.
- Chowdhury, M., Campbell, S. and Buckley, M. (2021). Proteomic analysis of archaeological ceramics from Tell Khaiber, southern Iraq. *Journal of archaeological science*, 132, p.105414. [Online]. Available at: doi:10.1016/j.jas.2021.105414.
- Cleland, T. P. (2018). Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *Journal of proteome research*, 17 (11), pp.3976–3983. [Online]. Available

at: doi:10.1021/acs.jproteome.8b00637.

Codron, D. et al. (2012). The confounding effects of source isotopic heterogeneity on consumer-diet and tissue-tissue stable isotope relationships. *Oecologia*, 169 (4), Springer Science and Business Media LLC., pp.939–953. [Online]. Available at: doi:10.1007/s00442-012-2274-3 [Accessed 19 October 2024].

Collins, Buckley and Grundy. (2010). ZooMS: the collagen barcode and fingerprints. *Spectroscopy Europe*, 22 (2), pp.6–10. [Online]. Available at: <https://www.spectroscopyeurope.com/article/zooms-collagen-barcode-and-fingerprints>.

Colonese, A. C. et al. (2017). New criteria for the molecular identification of cereal grains associated with archaeological artefacts. *Scientific reports*, 7 (1), p.6633. [Online]. Available at: doi:10.1038/s41598-017-06390-x.

Copley, M. S. et al. (2003). Direct chemical evidence for widespread dairying in prehistoric Britain. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (4), Proceedings of the National Academy of Sciences., pp.1524–1529. [Online]. Available at: doi:10.1073/pnas.0335955100.

Copley, M. S. et al. (2004). Organic residue evidence for the processing of marine animal products in pottery vessels from the pre-colonial archaeological site of Kasteelberg D east, South Africa. *South African Journal of Science*, 100 (5), journals.co.za., pp.279–283. [Online]. Available at: doi:10.10520/ejc96255.

Cottrell, J. S. (2011). Protein identification using MS/MS data. *Journal of proteomics*, 74 (10), Elsevier BV., pp.1842–1851. [Online]. Available at: doi:10.1016/j.jprot.2011.05.014 [Accessed 22 October 2024].

Courel, B. et al. (2020). Organic residue analysis shows sub-regional patterns in the use of pottery by Northern European hunter–gatherers. *Royal Society Open Science*, 7 (4), Royal Society., p.192016. [Online]. Available at: doi:10.1098/rsos.192016.

Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26 (12), Springer Science and Business Media LLC., pp.1367–1372. [Online]. Available at: doi:10.1038/nbt.1511 [Accessed 22 October 2024].

Craig, O. and Collins, M. J. (2000). An improved method for the immunological detection of mineral bound protein using hydrofluoric acid and direct capture. *Journal of immunological methods*, 236 (1-2), pp.89–97. [Online]. Available at: doi:10.1016/S0022-1759(99)00242-2.

Craig, O. E. (2004). Organic analysis of «food crusts» from sites in the Schelde valley, Belgium: a preliminary evaluation. *Notae Praehistoricae*, 24, pp.209–217. [Online]. Available at: https://biblio.naturalsciences.be/associated_publications/notae-praehistoricae/NP24/np24_218-219.pdf [Accessed 18 October 2024].

Craig, O. E. et al. (2005). Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity*, cambridge.org. [Online]. Available at: <https://www.cambridge.org/core/journals/antiquity/article/did-the-first-farmers-of-central-and-eastern-europe-produce-dairy-foods/284138196CFD83FA06340C061EDF5F93>.

Craig, O. E. et al. (2007). Molecular and isotopic demonstration of the processing of aquatic products in northern European prehistoric pottery. *Archaeometry*, 49 (1), Wiley., pp.135–152. [Online]. Available at: doi:10.1111/j.1475-4754.2007.00292.x [Accessed 19 October 2024].

- Craig, O. E. et al. (2013). Earliest evidence for the use of pottery. *Nature*, Nature Publishing Group. [Online]. Available at: doi:10.1038/nature12109.
- Craig, O. E. and Collins, M. J. (2002). The Removal of Protein from Mineral Surfaces: Implications for Residue Analysis of Archaeological Materials. *Journal of archaeological science*, 29 (10), pp.1077–1082. [Online]. Available at: doi:10.1006/jasc.2001.0757.
- Cramp, L. J. E. et al. (2014). Neolithic dairy farming at the extreme of agriculture in northern Europe. *Proceedings. Biological sciences*, 281 (1791), The Royal Society., p.20140819. [Online]. Available at: doi:10.1098/rspb.2014.0819.
- Cubas, M. et al. (2020). Latitudinal gradient in dairy production with the introduction of farming in Atlantic Europe. *Nature communications*, 11 (1), p.2036. [Online]. Available at: doi:10.1038/s41467-020-15907-4.
- Cubas, M. et al. (2023). Organic Residue Analysis of Archaeological Pottery. In: *Reference Module in Social Sciences*. Elsevier. [Online]. Available at: doi:10.1016/B978-0-323-90799-6.00047-1.
- Dallongeville, S. et al. (2011). Dealing with the identification of protein species in ancient amphorae. *Analytical and bioanalytical chemistry*, 399 (9), pp.3053–3063. [Online]. Available at: doi:10.1007/s00216-010-4218-2.
- Demirci, Ö. et al. (2020). First lipid residue analysis of Early Neolithic pottery from Swifterbant (the Netherlands, ca. 4300–4000 BC). *Archaeological and anthropological sciences*, 12 (5), p.105. [Online]. Available at: doi:10.1007/s12520-020-01062-w.
- Demirci, Ö. et al. (2021). Lipid residue analysis on Swifterbant pottery (c. 5000–3800 cal BC) in the Lower Rhine-Meuse area (the Netherlands) and its implications for human-animal interactions in relation to the Neolithisation process. *Journal of Archaeological Science: Reports*, 36, p.102812. [Online]. Available at: doi:10.1016/j.jasrep.2021.102812.
- Desmond, A. et al. (2018). ZooMS identification of bone tools from the North African Later Stone Age. *Journal of archaeological science*, 98, Elsevier., pp.149–157. [Online]. Available at: doi:10.1016/j.jas.2018.08.012.
- Doerr, A. (2015). DIA mass spectrometry. *Nature methods*, 12 (1), Springer Science and Business Media LLC., pp.35–35. [Online]. Available at: doi:10.1038/nmeth.3234 [Accessed 22 October 2024].
- Doliente, J. E. et al. (2024). Alkylresorcinol detection and identification in archaeological pottery using ultra-high-performance liquid chromatography-quadrupole/Orbitrap mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 38 (15), Wiley., p.e9771. [Online]. Available at: doi:10.1002/rcm.9771.
- Drieu, L. et al. (2020). Is it possible to identify ancient wine production using biomolecular approaches? *STAR: Science & Technology of Archaeological Research*, 6 (1), Routledge., pp.16–29. [Online]. Available at: doi:10.1080/20548923.2020.1738728.
- Drieu, L. et al. (2024). A medium-throughput approach for improved taxonomic identification of lipids preserved in ancient pottery. *Archaeometry*, Wiley. [Online]. Available at: doi:10.1111/arcm.12976.
- Eliuk, S. and Makarov, A. (2015). Evolution of Orbitrap mass spectrometry instrumentation. *Annual review of analytical chemistry (Palo Alto, Calif.)*, 8 (1), Annual Reviews., pp.61–80. [Online]. Available at: doi:10.1146/annurev-anchem-071114-040325 [Accessed 22 October

2024].

Evans, M. et al. (2024). The impact of cooking and burial on proteins: a characterisation of experimental foodcrusts and ceramics. *Royal Society open science*, 11 (240610). [Online]. Available at: doi:10.1098/rsos.240610.

Evershed, R. (2008). Organic residue analysis in archaeology: The archaeological biomarker revolution. *Archaeometry*, 50, repository.cam.ac.uk., pp.895–924. [Online]. Available at: doi:10.1111/J.1475-4754.2008.00446.X.

Evershed, R. P. et al. (2008). Experimental evidence for the processing of marine animal products and other commodities containing polyunsaturated fatty acids in pottery vessels. *Archaeometry*, 50 (1), Wiley., pp.101–113. [Online]. Available at: doi:10.1111/j.1475-4754.2007.00368.x.

Evershed, R. P. and Tuross, N. (1996). Proteinaceous Material from Potsherds and Associated Soils. *Journal of archaeological science*, 23 (3), pp.429–436. [Online]. Available at: doi:10.1006/jasc.1996.0038.

Fagnäs, Z. et al. (2024). Digging deeper into ancient skeletal proteomes through consecutive digestion with multiple proteases. *Journal of proteomics*, 298, p.105143. [Online]. Available at: doi:10.1016/j.jprot.2024.105143.

Fernandes, R. et al. (2014). Food reconstruction using isotopic transferred signals (FRUITS): a Bayesian model for diet reconstruction. *PloS one*, 9 (2), Public Library of Science (PLoS)., p.e87436. [Online]. Available at: doi:10.1371/journal.pone.0087436 [Accessed 19 October 2024].

Fischer, R. and Kessler, B. M. (2015). Gel-aided sample preparation (GASP)--a simplified method for gel-assisted proteomic sample generation from protein extracts and intact cells. *Proteomics*, 15 (7), Wiley., pp.1224–1229. [Online]. Available at: doi:10.1002/pmic.201400436.

Fonseca, B., Freeman, C. L. and Collins, M. J. (2022). Conformational analysis and water dynamics: a molecular dynamics study on the survival of a β -lactoglobulin peptide in the archaeological record. *Chemical physics*, 561 (111602), Elsevier BV., p.111602. [Online]. Available at: doi:10.1016/j.chemphys.2022.111602 [Accessed 11 June 2022].

Fonseca, B., Freeman, C. L. and Collins, M. J. (2024). The survival of β -lactoglobulin peptides in the archaeological record: Vulnerability vs. Sequence variation. *bioRxiv*, p.2024.09.13.612646. [Online]. Available at: doi:10.1101/2024.09.13.612646 [Accessed 17 September 2024].

Frahm, E. (2014). Scanning electron microscopy (SEM): Applications in archaeology. *Encyclopedia of global archaeology*, academia.edu. [Online]. Available at: https://www.academia.edu/download/33700829/Frahm_2014_SEM_in_Archaeology.pdf.

Ge, W. et al. (2024). Beyond rice: Preliminary investigation of encrusted organic residues in a pottery vessel reveals diverse food resources at the Neolithic Hemudu Settlement, China. *Journal of archaeological science, reports*, 58 (104714), Elsevier BV., p.104714. [Online]. Available at: doi:10.1016/j.jasrep.2024.104714.

Gobom, J. et al. (1999). Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *Journal of mass spectrometry*, 34 (2), Wiley., pp.105–116. [Online]. Available at:

doi:10.1002/(SICI)1096-9888(199902)34:2<105::AID-JMS768>3.0.CO;2-4.

González Carretero, L., Wollstonecroft, M. and Fuller, D. Q. (2017). A methodological approach to the study of archaeological cereal meals: a case study at Çatalhöyük East (Turkey). *Vegetation history and archaeobotany*, 26 (4), Springer Science and Business Media LLC., pp.415–432. [Online]. Available at: doi:10.1007/s00334-017-0602-6.

Graff, S. R. (2020). Archaeology of cuisine and cooking. *Annual review of anthropology*, 49 (1), Annual Reviews., pp.337–354. [Online]. Available at: doi:10.1146/annurev-anthro-102317-045734.

Granholm, V. and Käll, L. (2011). Quality assessments of peptide-spectrum matches in shotgun proteomics. *Proteomics*, 11 (6), Wiley., pp.1086–1093. [Online]. Available at: doi:10.1002/pmic.201000432 [Accessed 22 October 2024].

Gron, K. J., Montgomery, J. and Rowley-Conwy, P. (2015). Cattle management for dairying in Scandinavia's earliest Neolithic. *PloS one*, 10 (7), Public Library of Science (PLoS)., p.e0131267. [Online]. Available at: doi:10.1371/journal.pone.0131267.

Gu, M. and Buckley, M. (2018). Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. *BMC bioinformatics*, 19 (1), p.241. [Online]. Available at: doi:10.1186/s12859-018-2221-3.

Gupta, N. and Pevzner, P. A. (2009). False discovery rates of protein identifications: a strike against the two-peptide rule. *Journal of proteome research*, 8 (9), American Chemical Society (ACS)., pp.4173–4181. [Online]. Available at: doi:10.1021/pr9004794.

Gurdeep Singh, R. et al. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of proteome research*, 18 (2), pp.606–615. [Online]. Available at: doi:10.1021/acs.jproteome.8b00716.

Hansen, J. et al. (2024). A comparative study of commercially available, minimally invasive, sampling methods on Early Neolithic humeri analysed via palaeoproteomics. *Journal of archaeological science*, 167, p.106002. [Online]. Available at: doi:10.1016/j.jas.2024.106002.

Hansson, A.-M. and Isaksson, S. (1994). Analyses of charred organic remains. *Laborativ arkeologi*, 7, archaeology.su.se., pp.21–29. [Online]. Available at: https://www.archaeology.su.se/polopoly_fs/1.400583.1536668049!/menu/standard/file/LA7.Hansson.Isaksson.pdf.

Hao, Z. et al. (2012). A Quadrupole-Orbitrap Hybrid Mass Spectrometer Offers Highest Benchtop Performance for In-Depth Analysis of Complex Proteomes. *Thermo Scientific Application Notes*, 552. [Online]. Available at: <https://static.thermoscientific.com/images/D21364~.pdf>.

Heiss, A. G. et al. (2020). Mashies to Mashies, Crust to Crust. Presenting a novel microstructural marker for malting in the archaeological record. *PloS one*, 15 (5), p.e0231696. [Online]. Available at: doi:10.1371/journal.pone.0231696.

Hendy, J. et al. (2018a). A guide to ancient protein studies. *Nature ecology & evolution*, 2 (5), nature.com., pp.791–799. [Online]. Available at: doi:10.1038/s41559-018-0510-x.

Hendy, J. et al. (2018b). Proteomic evidence of dietary sources in ancient dental calculus. *Proceedings. Biological sciences / The Royal Society*, 285 (1883), royalsocietypublishing.org. [Online]. Available at: doi:10.1098/rspb.2018.0977.

- Heron, C. et al. (2015). Cooking fish and drinking milk? Patterns in pottery use in the southeastern Baltic, 3300–2400 cal BC. *Journal of archaeological science*, 63, Elsevier., pp.33–43. [Online]. Available at: doi:10.1016/j.jas.2015.08.002.
- Heron, C. et al. (2016). First molecular and isotopic evidence of millet processing in prehistoric pottery vessels. *Scientific reports*, 6, p.38767. [Online]. Available at: doi:10.1038/srep38767.
- Heron, C. and Craig, O. (2015). Aquatic Resources in Foodcrusts: Identification and Implication. *Radiocarbon*, 57 (4), Cambridge University Press., pp.707–719. [Online]. Available at: doi:10.2458/azu_rc.57.18454 [Accessed 12 May 2022].
- Heron, C. and Evershed, R. P. (1993). The analysis of organic residues and the study of pottery use. *Archaeological method and theory*, JSTOR. [Online]. Available at: https://www.jstor.org/stable/20170233?casa_token=yTiAlQm80ggAAAAA:tsY_ezNzdKT2Wp0-Zt_n6hxSMVa99awRsAYX2DaoOvxPZqPs_ohvhF_82pQdsFPQdFjm2n0FOsdFGvbM_dgouuCCHHhgi10Re8lIn_Zg_gXxmWDSZ-H71A.
- Hickinbotham, S. et al. (2020). How to get your goat: automated identification of species from MALDI-ToF spectra. *Bioinformatics*, 36 (12), pp.3719–3725. [Online]. Available at: <https://academic.oup.com/bioinformatics/article-abstract/36/12/3719/5807605>.
- Hughes, C. S. et al. (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology*, 10 (10), EMBO., p.757. [Online]. Available at: doi:10.15252/msb.20145625 [Accessed 21 October 2024].
- Hughes, C. S. et al. (2019). Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature protocols*, 14 (1), nature.com., pp.68–85. [Online]. Available at: doi:10.1038/s41596-018-0082-x.
- Hu, L. et al. (2017). Proteomic study of the effect of different cooking methods on protein oxidation in fish fillets. *RSC advances*, 7 (44), The Royal Society of Chemistry., pp.27496–27505. [Online]. Available at: doi:10.1039/C7RA03408C [Accessed 12 November 2021].
- Hu, Q. et al. (2005). The Orbitrap: a new mass spectrometer. *Journal of mass spectrometry*, 40 (4), Wiley., pp.430–443. [Online]. Available at: doi:10.1002/jms.856.
- Jacob, J. et al. (2008). Millet cultivation history in the French Alps as evidenced by a sedimentary molecule. *Journal of archaeological science*, 35 (3), Elsevier BV., pp.814–820. [Online]. Available at: doi:10.1016/j.jas.2007.06.006 [Accessed 18 October 2024].
- Jeong, K., Kim, S. and Bandeira, N. (2012). False discovery rates in spectral identification. *BMC bioinformatics*, 13 Suppl 16 (S16), Springer Science and Business Media LLC., p.S2. [Online]. Available at: doi:10.1186/1471-2105-13-S16-S2 [Accessed 16 October 2024].
- Keil-Dlouha, V. et al. (1971). Proteolytic activity of pseudotrypsin. *FEBS letters*, Elsevier. [Online]. Available at: <https://www.sciencedirect.com/science/article/pii/0014579371803733>.
- Keller, A. et al. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical chemistry*, 74 (20), American Chemical Society (ACS)., pp.5383–5392. [Online]. Available at: doi:10.1021/ac025747h [Accessed 22 October 2024].
- Kong, A. T. et al. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature methods*, 14 (5), Springer Science and

Business Media LLC., pp.513–520. [Online]. Available at: doi:10.1038/nmeth.4256.

Kubiak-Martens, L. (1999). The plant food component of the diet at the late Mesolithic (Ertebølle) settlement at Tybrind Vig, Denmark. *Vegetation history and archaeobotany*, 8 (1-2), Springer Science and Business Media LLC., pp.117–127. [Online]. Available at: doi:10.1007/bf02042850.

Kubiak-Martens, L. (2002). New evidence for the use of root foods in pre-agrarian subsistence recovered from the late Mesolithic site at Halsskov, Denmark. *Vegetation history and archaeobotany*, 11 (1-2), Springer Science and Business Media LLC., pp.23–32. [Online]. Available at: doi:10.1007/s003340200003.

Kubiak-Martens, L. (2016). Scanning electron microscopy and starchy food in Mesolithic Europe: the importance of roots and tubers in Mesolithic diet. In: Hardy, K. and Kubiak-Martens, L. (Eds). *Wild Harvest. Plants in the Hominin and pre-Agrarian Human Worlds*. Oxbow Books Oxford. pp.113–134.

Kubiak-Martens, L. et al. (2024). Transformation of cereal grains: Botanical and chemical analysis of food residues encrusted on pottery from the Funnel Beaker settlement of Oldenburg LA 77, northern Germany. *PloS one*, 19 (1), p.e0296986. [Online]. Available at: doi:10.1371/journal.pone.0296986.

Kubiak-Martens, L., Brinkkemper, O. and Oudemans, T. F. M. (2015). What's for dinner? Processed food in the coastal area of the northern Netherlands in the Late Neolithic. *Vegetation history and archaeobotany*, 24 (1), Springer Science and Business Media LLC., pp.47–62. [Online]. Available at: doi:10.1007/s00334-014-0485-8.

Lanigan, L. T. et al. (2020). Multi-protease analysis of Pleistocene bone proteomes. *Journal of proteomics*, 228, Elsevier., p.103889. [Online]. Available at: doi:10.1016/j.jprot.2020.103889.

Le Meillour, L. et al. (2024). Increasing sustainability in palaeoproteomics by optimising digestion times for large-scale archaeological bone analyses. *iScience*, cell.com. [Online]. Available at: [https://www.cell.com/iscience/pdf/S2589-0042\(24\)00653-9.pdf](https://www.cell.com/iscience/pdf/S2589-0042(24)00653-9.pdf).

Lucquin, A. et al. (2016). Ancient lipids document continuity in the use of early hunter–gatherer pottery through 9,000 years of Japanese prehistory. *Proceedings of the National Academy of Sciences*, 113 (15), pp.3991–3996. [Online]. Available at: doi:10.1073/pnas.1522908113.

Lundy, J. et al. (2023). Cuisine in transition? Organic residue analysis of domestic containers from 9th–14th century Sicily. *Royal Society open science*, 10 (3), p.221305. [Online]. Available at: doi:10.1098/rsos.221305.

Lu, X. and Zhu, H. (2005). Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Molecular & cellular proteomics: MCP*, 4 (12), Elsevier., pp.1948–1958. [Online]. Available at: doi:10.1074/mcp.M500138-MCP200 [Accessed 21 October 2024].

Lyu, N. et al. (2024). Characterization of Pottery Foodcrusts Through Lipid and Proteomic Analyses: A Case Study from the Xiawan Site in Yixing City, East China. *Journal of archaeological science*, 161 (105902), papers.ssrn.com. [Online]. Available at: doi:10.1016/j.jas.2023.105902 [Accessed 8 September 2023].

Ma, B. (2015). Novor: real-time peptide de novo sequencing software. *Journal of the American Society for Mass Spectrometry*, 26 (11), American Chemical Society (ACS).,

pp.1885–1894. [Online]. Available at: doi:10.1007/s13361-015-1204-0 [Accessed 22 October 2024].

Mackie, M. et al. (2018). Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angewandte Chemie*, 57 (25), pp.7369–7374. [Online]. Available at: doi:10.1002/anie.201713020.

Ma, M. et al. (2024). Forager-farmer transition at the crossroads of East and Southeast Asia 4900 years ago. *Science bulletin*, 69 (1), Elsevier BV., pp.103–113. [Online]. Available at: doi:10.1016/j.scib.2023.10.015 [Accessed 28 October 2024].

Marcotte, E. M. (2007). How do shotgun proteomics algorithms identify proteins? *Nature biotechnology*, 25 (7), Springer Science and Business Media LLC., pp.755–757. [Online]. Available at: doi:10.1038/nbt0707-755.

Martins, C. P. B. et al. (2016). Orbitrap Mass Spectrometry. In: *Applications of Time-of-Flight and Orbitrap Mass Spectrometry in Environmental, Food, Doping, and Forensic Analysis*. Comprehensive analytical chemistry. 71. Elsevier. pp.3–18. [Online]. Available at: doi:10.1016/bs.coac.2016.01.001 [Accessed 22 October 2024].

McGrath, K. et al. (2019). Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points. *Scientific reports*, 9 (1), nature.com., p.11027. [Online]. Available at: doi:10.1038/s41598-019-47299-x.

Meier, F. et al. (2018). Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Molecular & cellular proteomics: MCP*, 17 (12), Elsevier BV., pp.2534–2545. [Online]. Available at: doi:10.1074/mcp.TIR118.000900.

Melby, J. A. et al. (2021). Novel strategies to address the challenges in top-down proteomics. *Journal of the American Society for Mass Spectrometry*, 32 (6), American Chemical Society (ACS)., pp.1278–1294. [Online]. Available at: doi:10.1021/jasms.1c00099.

Michelmann, K. et al. (2015). Fundamentals of trapped ion mobility spectrometry. *Journal of the American Society for Mass Spectrometry*, 26 (1), American Chemical Society (ACS)., pp.14–24. [Online]. Available at: doi:10.1007/s13361-014-0999-4.

Mikesh, L. M. et al. (2006). The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et biophysica acta*, 1764 (12), Elsevier BV., pp.1811–1822. [Online]. Available at: doi:10.1016/j.bbapap.2006.10.003.

Miller, M. J. et al. (2020). Interpreting ancient food practices: stable isotope and molecular analyses of visible and absorbed residues from a year-long cooking experiment. *Scientific reports*, 10 (1), p.13704. [Online]. Available at: doi:10.1038/s41598-020-70109-8.

Mooradian, A. D. et al. (2020). ProteoClade: A taxonomic toolkit for multi-species and metaproteomic analysis. *PLoS computational biology*, 16 (3), p.e1007741. [Online]. Available at: doi:10.1371/journal.pcbi.1007741.

Müller, T. and Winter, D. (2017). Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. *Molecular & cellular proteomics: MCP*, 16 (7), Elsevier., pp.1173–1187. [Online]. Available at: doi:10.1074/mcp.M116.064048 [Accessed 20 October 2024].

Nielsen-Marsh, C. M. et al. (2002). Sequence preservation of osteocalcin protein and mitochondrial DNA in bison bones older than 55 ka. *Geology*, 30 (12), Geological Society of America., p.1099. [Online]. Available at:

doi:10.1130/0091-7613(2002)030<1099:spoop>2.0.co;2.

Olsen, J. V. et al. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods*, 4 (9), Springer Science and Business Media LLC., pp.709–712. [Online]. Available at: doi:10.1038/nmeth1060.

Oras, E. et al. (2017). The adoption of pottery by north-east European hunter-gatherers: Evidence from lipid residue analysis. *Journal of archaeological science*, 78, pp.112–119. [Online]. Available at: doi:10.1016/j.jas.2016.11.010.

Ostrom, P. H. et al. (2006). Unraveling the sequence and structure of the protein osteocalcin from a 42ka fossil horse. *Geochimica et cosmochimica acta*, 70 (8), Elsevier BV., pp.2034–2044. [Online]. Available at: doi:10.1016/j.gca.2006.01.004 [Accessed 22 October 2024].

Papakosta, V., Oras, E. and Isaksson, S. (2019). Early pottery use across the Baltic—a comparative lipid residue study on Ertebølle and Narva ceramics from coastal hunter-gatherer sites in southern Scandinavia, northern Germany and Estonia. *Journal of archaeological science*, 24, Elsevier., pp.142–151. [Online]. Available at: doi:10.1016/j.jasrep.2019.01.003.

Parker, G. J. et al. (2019). Sex estimation using sexually dimorphic amelogenin protein fragments in human enamel. *Journal of archaeological science*, 101, Elsevier., pp.169–180. [Online]. Available at: doi:10.1016/j.jas.2018.08.011.

Perkins, D. N. et al. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20 (18), Wiley., pp.3551–3567. [Online]. Available at: doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Phillips, D. L. and Gregg, J. W. (2003). Source partitioning using stable isotopes: coping with too many sources. *Oecologia*, 136 (2), Springer Science and Business Media LLC., pp.261–269. [Online]. Available at: doi:10.1007/s00442-003-1218-3 [Accessed 19 October 2024].

Pino, L. K. et al. (2020). Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries. *Molecular & cellular proteomics: MCP*, 19 (7), Elsevier BV., pp.1088–1103. [Online]. Available at: doi:10.1074/mcp.P119.001913.

Presslee, S. (2020). *ZooMS Markers: Published Data*. [Online]. Available at: <https://docs.google.com/spreadsheets/d/1ipm9fFFyha8IEzRO2F5zVXIkl0ldwYiWgX5pGqETzBco/edit> [Accessed 3 April 2023].

Qi, Y. and Volmer, D. A. (2017). Electron-based fragmentation methods in mass spectrometry: An overview. *Mass spectrometry reviews*, 36 (1), Wiley Online Library., pp.4–15. [Online]. Available at: doi:10.1002/mas.21482.

Ramsøe, A. et al. (2020). DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *Journal of archaeological science*, 115, p.105080. [Online]. Available at: doi:10.1016/j.jas.2020.105080.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*, 2 (8), pp.1896–1906. [Online]. Available at: doi:10.1038/nprot.2007.261.

Robinson, N. E. and Robinson, A. B. (2001). Molecular clocks. *Proceedings of the National*

Academy of Sciences of the United States of America, 98 (3), Proceedings of the National Academy of Sciences., pp.944–949. [Online]. Available at: doi:10.1073/pnas.98.3.944 [Accessed 22 October 2024].

Robson, H. K. et al. (2020). Walnuts, salmon and sika deer: Exploring the evolution and diversification of Jōmon ‘culinary’ traditions in prehistoric Hokkaidō. *Journal of Anthropological Archaeology*, 60, p.101225. [Online]. Available at: doi:10.1016/j.jaa.2020.101225.

Robson, H. K. et al. (2021). Organic residue analysis of Early Neolithic ‘bog pots’ from Denmark demonstrates the processing of wild and domestic foodstuffs. *Journal of Archaeological Science: Reports*, 36, Elsevier., p.102829. [Online]. Available at: doi:10.1016/j.jasrep.2021.102829.

Robson, H. K. et al. (2022). Light production by ceramic using Hunter-gatherer-fishers of the circum-Baltic. *Proceedings of the Prehistoric Society*, 88, Cambridge University Press (CUP)., pp.25–52. [Online]. Available at: doi:10.1017/ppr.2022.12 [Accessed 17 November 2022].

Rodriguez, J. et al. (2008). Does trypsin cut before proline? *Journal of proteome research*, 7 (1), American Chemical Society (ACS)., pp.300–305. [Online]. Available at: doi:10.1021/pr0705035.

Rodriguez Palomo, I. et al. (2024). Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *Peer Community J.*, (e107), p.2023.12.15.571577. [Online]. Available at: doi:10.24072/pcjournal.491 [Accessed 29 November 2024].

Rüther, P. L. et al. (2022). SPIN enables high throughput species identification of archaeological bone by proteomics. *Nature communications*, 13 (1), nature.com., p.2458. [Online]. Available at: doi:10.1038/s41467-022-30097-x.

Salque, M. et al. (2013). Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature*, 493 (7433), Springer Science and Business Media LLC., pp.522–525. [Online]. Available at: doi:10.1038/nature11698 [Accessed 18 July 2024].

Schroeter, E. R. and Cleland, T. P. (2016). Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid communications in mass spectrometry: RCM*, 30 (2), Wiley Online Library., pp.251–255. [Online]. Available at: doi:10.1002/rcm.7445.

Sechi, S. and Chait, B. T. (1998). Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Analytical chemistry*, 70 (24), American Chemical Society (ACS)., pp.5150–5158. [Online]. Available at: doi:10.1021/ac9806005.

Segerberg, A. et al. (1991). Ceramic chronology in view of ¹⁴C datings. *Laborativ arkeologi*, 5, pp.83–91. [Online]. Available at: https://www.su.se/polopoly_fs/1.400489.1536662261!/menu/standard/file/Lab%20ark%205.Segerberg%2C%20Possnert%2C%20Arrhenius%20%26%20Lid%20C3%A9n.pdf.

Shevchenko, A. et al. (2018). Fine Endmesolithic fish caviar meal discovered by proteomics in foodcrusts from archaeological site Friesack 4 (Brandenburg, Germany). *PloS one*, 13 (11), p.e0206483. [Online]. Available at: doi:10.1371/journal.pone.0206483.

Shoda, S. et al. (2017). Pottery use by early Holocene hunter-gatherers of the Korean peninsula closely linked with the exploitation of marine resources. *Quaternary Science Reviews*, 170, pp.164–173. [Online]. Available at: doi:10.1016/j.quascirev.2017.06.032.

Sobott, F. et al. (2009). Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination. *Journal of the American Society for Mass Spectrometry*, 20 (9), Springer., pp.1652–1659. [Online]. Available at: doi:10.1016/j.jasms.2009.04.023.

Solazzo, C. et al. (2008). Identification of protein remains in archaeological potsherds by proteomics. *Analytical chemistry*, 80 (12), pp.4590–4597. [Online]. Available at: doi:10.1021/ac800515v.

Solntsev, S. K. et al. (2018). Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of proteome research*, 17 (5), ACS Publications., pp.1844–1851. [Online]. Available at: doi:10.1021/acs.jproteome.7b00873.

Sørensen, L. and Karg, S. (2014). The expansion of agrarian societies towards the north – new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. *Journal of archaeological science*, 51, Elsevier BV., pp.98–114. [Online]. Available at: doi:10.1016/j.jas.2012.08.042.

Strohalm, M. et al. (2010). mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Analytical chemistry*, 82 (11), ACS Publications., pp.4648–4651. [Online]. Available at: doi:10.1021/ac100818g.

Styring, A. K. et al. (2013). The effect of charring and burial on the biochemical composition of cereal grains: investigating the integrity of archaeological plant material. *Journal of archaeological science*, 40 (12), Elsevier BV., pp.4767–4779. [Online]. Available at: doi:10.1016/j.jas.2013.03.024.

Szpak, P. (2011). Fish bone chemistry and ultrastructure: implications for taphonomy and stable isotope analysis. *Journal of archaeological science*, 38 (12), Elsevier BV., pp.3358–3372. [Online]. Available at: doi:10.1016/j.jas.2011.07.022 [Accessed 31 October 2024].

Teetaert, D. et al. (2024). Pottery use across the Neolithic transition in northern Belgium: evidence from isotopic, molecular and microscopic analysis. *Archaeological and anthropological sciences*, 16 (8), Springer Science and Business Media LLC. [Online]. Available at: doi:10.1007/s12520-024-02030-4.

Thiede, B. et al. (2005). Peptide mass fingerprinting. *Methods (San Diego, Calif.)*, 35 (3), Elsevier BV., pp.237–247. [Online]. Available at: doi:10.1016/j.ymeth.2004.08.015 [Accessed 22 October 2024].

Trampota, F. and Kvěťina, P. (2020). How do they fit together? A case study of Neolithic pottery typology and radiocarbon chronology. *Archivaria*, 72, archeologickerozhledy.cz., pp.163–193. [Online]. Available at: doi:10.35686/ar.2020.6.

Vandermarliere, E., Mueller, M. and Martens, L. (2013). Getting intimate with trypsin, the leading protease in proteomics: TRYPSIN IN PROTEOMICS. *Mass Spectrometry Reviews*, 32 (6), Wiley Online Library., pp.453–465. [Online]. Available at: doi:10.1002/mas.21376.

Végh, E. I. and Douka, K. (2024). SpecieScan: Semi-Automated taxonomic identification of bone collagen peptides from MALDI-ToF MS spectra. *Bioinformatics* . [Online]. Available at: doi:10.1093/bioinformatics/btae054.

da Veiga Leprevost, F. et al. (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nature methods*, 17 (9), Springer Science and Business Media LLC., pp.869–870. [Online]. Available at: doi:10.1038/s41592-020-0912-y [Accessed 22 October 2024].

- Warinner, C., Korzow Richter, K. and Collins, M. J. (2022). Paleoproteomics. *Chemical reviews*, ACS Publications. [Online]. Available at: doi:10.1021/acs.chemrev.1c00703.
- Welker, F. et al. (2015). Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature*, 522 (7554), nature.com., pp.81–84. [Online]. Available at: doi:10.1038/nature14249.
- Welker, F. et al. (2020). The dental proteome of Homo antecessor. *Nature*, 580 (7802), pp.235–238. [Online]. Available at: doi:10.1038/s41586-020-2153-8.
- Wenger, C. D. and Coon, J. J. (2013). A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *Journal of proteome research*, 12 (3), ACS Publications., pp.1377–1386. [Online]. Available at: doi:10.1021/pr301024c.
- Wilkin, S. et al. (2024). Sequential trypsin and ProAlanase digestions unearth immunological protein biomarkers shrouded by skeletal collagen. *iScience*, 27 (5), Elsevier BV., p.109663. [Online]. Available at: doi:10.1016/j.isci.2024.109663 [Accessed 21 October 2024].
- Wilson, J., van Doorn, N. L. and Collins, M. J. (2012). Assessing the extent of bone degradation using glutamine deamidation in collagen. *Analytical chemistry*, 84 (21), pp.9041–9048. [Online]. Available at: doi:10.1021/ac301333t.
- Wiśniewski, J. R. et al. (2009). Universal sample preparation method for proteome analysis. *Nature methods*, 6 (5), nature.com., pp.359–362. [Online]. Available at: doi:10.1038/nmeth.1322.
- Wiśniewski, J. R. (2019). Filter Aided Sample Preparation - A tutorial. *Analytica chimica acta*, 1090, Elsevier BV., pp.23–30. [Online]. Available at: doi:10.1016/j.aca.2019.08.032.
- Yilmaz, M. et al. (2024). Sequence-to-sequence translation from mass spectra to peptides with a transformer model. *Nature communications*, 15 (1), Springer Science and Business Media LLC., p.6427. [Online]. Available at: doi:10.1038/s41467-024-49731-x [Accessed 6 September 2024].
- Yoneda, M. et al. (2019). Interpretation of bulk nitrogen and carbon isotopes in archaeological foodcrusts on potsherds. *Rapid communications in mass spectrometry: RCM*, 33 (12), Wiley Online Library., pp.1097–1106. [Online]. Available at: doi:10.1002/rcm.8446.
- Yu, F. et al. (2020). Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Molecular & cellular proteomics: MCP*, 19 (9), Elsevier BV., pp.1575–1585. [Online]. Available at: doi:10.1074/mcp.TIR120.002048.
- Yu, F., Deng, Y. and Nesvizhskii, A. (2024). MSFragger-DDA+ Enhances Peptide Identification Sensitivity with Full Isolation Window Search. *bioRxiv*, p.2024.10.12.618041. [Online]. Available at: doi:10.1101/2024.10.12.618041 [Accessed 15 October 2024].
- Yu, F., Haynes, S. E. and Nesvizhskii, A. I. (2021). IonQuant enables accurate and sensitive label-free quantification with FDR-controlled match-between-runs. *Molecular & cellular proteomics: MCP*, 20 (100077), Elsevier BV., p.100077. [Online]. Available at: doi:10.1016/j.mcpro.2021.100077 [Accessed 22 October 2024].
- Yu, T.-Y. et al. (2017). Cooking-induced protein modifications in meat. *Comprehensive Reviews in Food Science and Food Safety*, 16 (1), Wiley., pp.141–159. [Online]. Available at: doi:10.1111/1541-4337.12243.
- Zhang, J. et al. (2012). PEAKS DB: de novo sequencing assisted database search for

sensitive and accurate peptide identification. *Molecular & cellular proteomics: MCP*, 11 (4), Elsevier BV., p.M111.010587. [Online]. Available at: doi:10.1074/mcp.M111.010587.

Zhang, N. and Li, L. (2004). Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid communications in mass spectrometry: RCM*, 18 (8), Wiley., pp.889–896. [Online]. Available at: doi:10.1002/rcm.1423.

Zhou, W. et al. (2006). Fundamentals of scanning electron microscopy (SEM). In: *Scanning Microscopy for Nanotechnology*. New York, NY : Springer New York. pp.1–40. [Online]. Available at: doi:10.1007/978-0-387-39620-0_1.

Zohar, I. et al. (2008). The living and the dead: How do taphonomic processes modify relative abundance and skeletal completeness of freshwater fish? *Palaeogeography, palaeoclimatology, palaeoecology*, 258 (4), Elsevier BV., pp.292–316. [Online]. Available at: doi:10.1016/j.palaeo.2007.11.004 [Accessed 31 October 2024].

2 Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands

Joannes A. A. Dekker^{*1,2}, Dorothea Mylopotamitaki^{3,4}, Annemieke Verbaas⁵, Virginie Sinet-Mathiot^{4,6,7}, Samantha Presslee¹, Morgan L. McCarthy², Morten Tange Olsen², Jesper V. Olsen⁸, Youri van den Hurk^{9,10}, Joris Brattinga¹¹, Frido Welker²

1. BioArCh, Department of Archaeology, University of York, York, UK
2. Globe Institute, University of Copenhagen, Copenhagen, Denmark
3. Chaire de Paléanthropologie, Collège de France, Paris, France
4. Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany.
5. Faculty of Archaeology, Universiteit Leiden, Leiden, the Netherlands.
6. Univ. Bordeaux, CNRS, Ministère de la Culture, PACEA, UMR 5199, Pessac, France
7. Univ. Bordeaux, CNRS, Bordeaux INP, CBMN, UMR 5248 and Bordeaux Proteome Platform, Bordeaux, France
8. Novo Nordisk Foundation Center for Protein Research, Center for Protein Research, University of Copenhagen, Copenhagen, Denmark.
9. Groningen Institute of Archaeology, University of Groningen, Groningen, Netherlands.
10. Department of Archaeology and Cultural History, Norwegian University of Science and Technology, Norway.
11. Archol bv, Leiden, the Netherlands.

Published article:

Dekker, J. et al. (2024). Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands. *Peer Community Journal* 4 (e81). Available at doi:10.24072/pcjournal.451.

Research article

Published
2024-08-30

Cite as

Joannes A. A. Dekker, Dorothea Mylopotamitaki, Annemieke Verbaas, Virginie Sinet-Mathiot, Samantha Presslee, Morgan L. McCarthy, Morten Tange Olsen, Jesper V. Olsen, Youri van den Hurk, Joris Brattinga and Frido Welker (2024) *Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands*, Peer Community Journal, 4: e81.

Correspondence

jan@palaeome.org

Peer-review

Peer reviewed and recommended by
PCI Archaeology,

<https://doi.org/10.24072/pci.archaeo.100501>



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

Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands

Joannes A. A. Dekker^{1,2}, Dorothea Mylopotamitaki^{3,4}, Annemieke Verbaas⁵, Virginie Sinet-Mathiot^{4,6,7}, Samantha Presslee^{1,8}, Morgan L. McCarthy^{1,2}, Morten Tange Olsen^{1,2}, Jesper V. Olsen^{1,8}, Youri van den Hurk^{1,9,10}, Joris Brattinga¹¹, and Frido Welker^{1,2}

Volume 4 (2024), article e81

<https://doi.org/10.24072/pcjournal.451>

Abstract

Identification of the taxonomic origin of bone tools is an important, but often complicated, component of studying past societies. The species used for bone tool production provide insight into what species were exploited, potentially how, and for what purpose. Additionally, the choice of species may have important implications for the place of the tool within the larger toolkit. However, the taxonomic identification of bone tools is often unsuccessful based on morphology. Here we apply three palaeoproteomic techniques, ZooMS, SPIN-like data analysis and a targeted database search to narrow down the taxonomic identification of an unusually large Bronze Age bone tool from Heiloo, the Netherlands, to the North Atlantic right whale (*Eubalaena glacialis*). Additionally, the tool was investigated for use-wear, which showed that it was likely used for the processing of plant fibres. The assignment of the tool as whale bone adds support to the exploitation of whales by coastal Bronze Age populations, not just for meat, as previously suggested, but also for bone as a resource for tool production. We know of no other parallel of a bone tool such as this in terms of size, use, hafting, and taxonomic identity.

¹BioArCh, Department of Archaeology, University of York, York, UK, ²Section for Molecular Ecology and Evolution, Globe Institute, University of Copenhagen, Copenhagen, Denmark, ³Chaire de Paléanthropologie, Collège de France, Paris, France, ⁴Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany, ⁵Faculty of Archaeology, Universiteit Leiden, Leiden, the Netherlands, ⁶Univ. Bordeaux, CNRS, Ministère de la Culture, PACEA, UMR 5199, Pessac, France, ⁷Univ. Bordeaux, CNRS, Bordeaux INP, CBMN, UMR 5248 and Bordeaux Proteome Platform, Bordeaux, France, ⁸Novo Nordisk Foundation Center for Protein Research, Center for Protein Research, University of Copenhagen, Copenhagen, Denmark, ⁹Groningen Institute of Archaeology, University of Groningen, Groningen, Netherlands, ¹⁰Department of Archaeology and Cultural History, Norwegian University of Science and Technology, Norway, ¹¹Archol bv, Leiden, the Netherlands



Introduction

A common framework for studying archaeological tools and their place in past societies is to reconstruct an artefact biography, which contains the conception, manufacture, use, potential reuse and deposition of the artefact. The biography approach is commonly applied to a large variety of artefacts, but in the case of osseous artefacts the first step, the selection of raw material, remains poorly explored. The difficulty of identifying the species of a bone tool based on morphological characteristics is largely to blame for this, since morphologically diagnostic features are often removed during the shaping and use of a bone artefact.

Several biomolecular techniques have been proposed to resolve this problem, such as general applications of ancient DNA (Hofreiter & Pacher, 1997; Essel et al., 2023) and shotgun proteomic analysis (Multari et al., 2023), as well as some more specialised protein-based techniques like ZooMS (Zooarchaeology by Mass Spectrometry) (Buckley et al., 2009) and SPIN (Species by Proteome INvestigation) (Rüther et al., 2022). Each of these techniques comes with their own specific advantages and disadvantages. Ancient DNA analysis is understood to be the most precise in its taxonomic identifications, but it requires a relatively high standard of biomolecular preservation. Additionally, it often requires relatively large sample sizes (Parker et al., 2021; Tejero et al., 2024), which may not be desirable for rare archaeological objects. However, it must be noted that recent developments on minimally invasive DNA extraction may have removed this obstacle (Essel et al., 2023; Tejero et al., 2024). On the other side of the spectrum, ZooMS can often only assign taxon up to genus level, but is able to handle less well preserved samples. SPIN and shotgun proteomics seem to require a roughly similar level of preservation as ZooMS, as all are targeting ancient proteins, however, they should be able to provide a more precise taxonomic identification, allowing a species level identification in cases where ZooMS can only provide a genus-level assignment. Conventional sample sizes for ZooMS and SPIN are relatively small, ranging from 5-20 mg (Welker et al., 2015; Sinet-Mathiot et al., 2019; Rüther et al., 2022), while for shotgun proteomics samples of 15-60 mg are more commonly taken (Procopio et al., 2017; Sawafuji et al., 2017; Cleland, 2018). The development of minimally invasive sampling protocols for palaeoproteomic analysis, primarily for ZooMS, is more advanced than for aDNA analysis. A variety of different sampling protocols have been proposed that seem promising, although their ability to extract ancient protein appears to be context-specific (McGrath et al., 2019; Evans et al., 2023; Hansen, Dekker, et al., 2024). In this study we chose to sample destructively for three reasons. First of all, the biomolecular preservation at the site was unknown. Secondly, the size of the object allowed for a sufficient sample to be extracted without compromising other analyses of the artefact, including visual ones. Moreover, use-wear analysis was performed before destructive sampling to prevent any interference following. The combination of use-wear analysis and proteomic analysis is becoming more common and should be considered as a 'best practice' (Bradfield et al., 2019; Dekker et al., 2021; Orłowska et al., 2023; Hansen, Sierra, et al., 2024). Considering that even some minimally invasive sampling protocols for ZooMS may obscure traces of use it is vital that use-wear analysis is performed in advance to prevent the loss of valuable information regarding the use of the tool (Sinet-Mathiot et al., 2021; Hansen, Dekker, et al., 2024). Lastly, the aim of this research was to refine the taxonomic identification to species level for which maximising sequence coverage is essential. It has been shown that destructive methods extract a larger abundance of protein than minimally invasive methods (Hansen, Dekker, et al., 2024).

Palaeoproteomic techniques, both minimally invasive and destructive, have been applied in a number of previous studies to study species selection in the production of osseous tools. These studies demonstrate how palaeoproteomic techniques can be used to reveal the use of unexpected species for bone tool production (Surovell et al., 2024), such as human bone (McGrath et al., 2019; Dekker et al., 2021), as well as the intentional selection of certain species (Desmond et al., 2018; Martisius et al., 2020; Bradfield et al., 2021; Adamczak et al., 2021).

In this study we apply multiple biomolecular techniques to a worked bone artefact (Specimen 2199) from the Bronze Age site of Heiloo-Zuiderloo, the Netherlands. Initial morphological inspection suggested that the bone artefact was of Elephantidae origin, which considering the lack of any Elephantidae species living in the Netherlands during the Bronze Age, would suggest either the use of (sub-)fossil mammoth bone or long-distance trade. To test this hypothesis we first analysed the bone artefact using ZooMS. Additionally, use-wear analysis was performed on the bone tool to study its production and use for a better understanding of the place of this artefact in the Bronze Age toolkit. However, ZooMS was not able to provide a species level identification for the bone tool, so another recently introduced palaeoproteomic technique, SPIN (Rüther et al., 2022), was performed in order to refine the taxonomic identification. The SPIN analysis revealed no collagen sequences were available for several of the species of interest. Thus, in order to resolve this blind spot, we acquired new collagen reference DNA sequences from selected cetacean species and performed a final targeted database search for a more precise taxonomic identification.

Materials and methods

Description of the site and context

The main subject of this study is a large plano-convex bone artefact (Specimen 2199). It was found in 2020 during archaeological excavations in Heiloo-Zuiderloo, The Netherlands (Figure 1). The site is located on the Dutch coastal area in the northwestern part of the country. The first traces of habitation in the area date from the transition period of the Neolithic to the Early Bronze Age (ca. 2000 BC). The preservation conditions of the prehistoric archaeological remains from the region are excellent due to a thick 2 metre layer of sand covering them. Additionally, both the archaeological remains and the sand layer are located below the groundwater level. As a result, many remarkable archaeological finds have been recovered in the area in recent years (van der Heiden, 2018; de Koning & Tuinman, 2019; Moesker et al., 2021; Brattinga, 2023).

During the 2020 excavations, a Bronze Age settlement along with adjacent agricultural fields were uncovered. These findings have improved our knowledge of the Bronze Age habitation of the region, showing that the settlements were located on higher parts of the coastal landscape, specifically the dunes, while the lower parts of the landscape were utilised for various activities such as agriculture, hunting, fishing, and livestock grazing (Brattinga, 2023). The faunal assemblage at the site is dominated by cattle (*Bos taurus*) and sheep/goat (*Ovis aries/Capra hircus*). Other domesticates such as pig (*Sus scrofa domesticus*), dog (*Canis familiaris*) and horse (*Equus caballus*) are present in smaller numbers, as are a number of wild taxa, such as whale (Cetacea), grey seal (*Halichoerus grypus*), red deer (*Cervus elaphus*) and the common murre (*Uria aalge*). However, each of these aforementioned taxa is represented by only one or two specimens. The site also yielded a number of fish remains, the majority of which was ascribed to cod (*Gadus morhua*). The bone assemblage at the site also features a small number of worked bone artefacts. Apart from the bone tool analysed in this study, two modified cattle scapulae were found, which have been interpreted as digging implements (Nieweg, 2023).

The bone object was found at the bottom of a Middle Bronze Age watering hole. Radiocarbon dates of the wooden shaft and botanical remains from the watering hole indicate the tool was made and used around 1500 BC (Laboratory ID: Poz-125900, 3230 ± 30 BP, 1541-1425 cal. BC, laboratory ID: Poz-156692, 3290 ± 35 BP, 1631-1456 cal BC, chronological ages calculated through OxCal 4.4.4 using the INTCAL20 calibration curve (Reimer et al., 2020)).

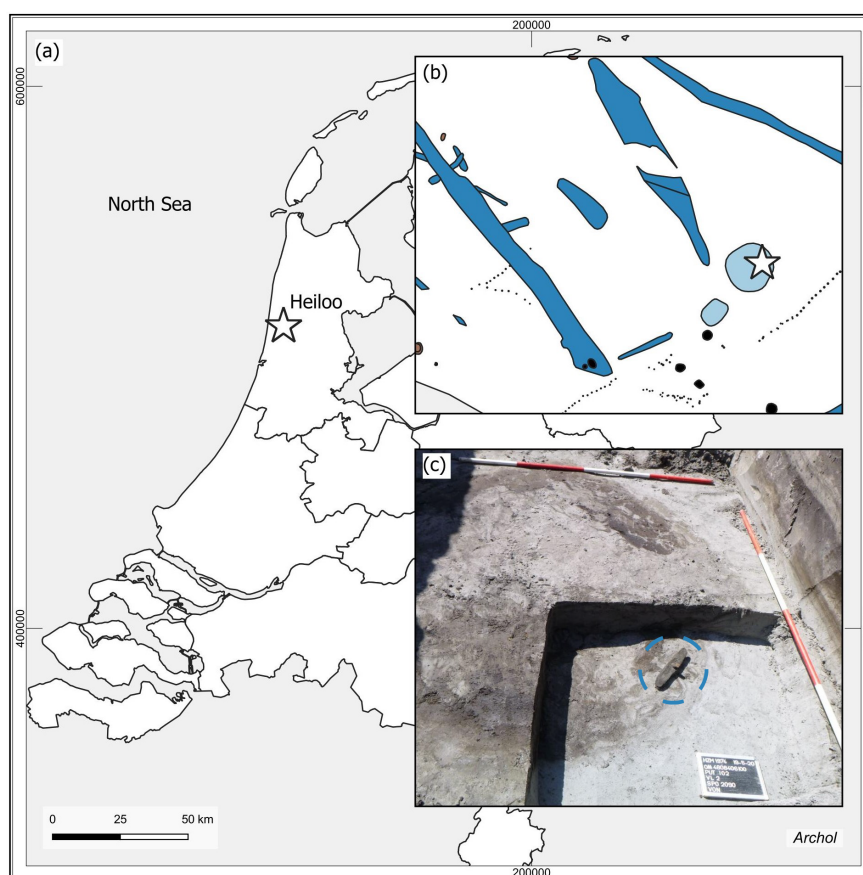


Figure 1 – a) Map of the Netherlands with a star indicating the site location. b) Part of the excavation plan. The find location of the analysed object is marked with a star. c) Photograph showing the bone tool in situ inside the round feature of the watering hole, as indicated by the blue circle. © Archol, Leiden.

Description of the artefact (Specimen 2199)

The artefact has a size of 33 x 6 x 3.5 cm and has a plano-convex shape with adze-like shaped ends (Figure 2). In the middle of the object, there is a square hole measuring 2 x 2 cm where a wooden handle has been inserted at an oblique angle. The inside shape of the hole causes the shaft to be fixed with a 50 degree angle. At the thickest point, the bone is 3.5 cm thick. The upper part has a smooth and shiny surface. The other side has an irregular texture with traces of the original spongy structure of the bone. The surface of the object shows elaborate use-wear traces.

The wooden handle was damaged during excavation, resulting in the distal part of it being lost. Therefore, the original length of the handle is unknown. After the artefact was recovered, it became clear that the handle was secured within the shaft of the tool by a wooden wedge. The wedge is inserted between the bone and the handle from the front and is made from the outer part of a willow branch. Based on the anatomical features of the cross, radial and tangential section it was determined that the handle itself is made from a modified oak branch (van Hees & van Amerongen, 2023). The branch is thickened 2 cm from the top, where a knob has been shaped using a sharp metal object. The finishing marks on the knob are relatively coarse and functional.

Three samples consisting of small bone flakes (28.6 mg, 21.0 mg and 10.7 mg) were taken from the artefact by scraping the bone with a scalpel. Two were taken from the inside of the shaft hole, the third from the downside of the artefact. The first sample was sent to BioArCh, University of York, the second to the Department for Evolutionary Anthropology of the Max Planck Institute in Leipzig, and the third sample was sent to the Globe Institute, University of Copenhagen. The sample to York was designated

for ZooMS analysis, whereas the sample sent to Leipzig was analysed with both ZooMS and LC-MS/MS. Lastly the sample sent to Copenhagen would be analysed using LC-MS/MS. ZooMS analysis was performed at York and Leipzig. Although unintended, an independent validation of the obtained MALDI-ToF MS spectra provides further validation of the resulting taxonomic identification of this unique bone artefact. From our knowledge, inter-laboratory comparisons are either rare in ZooMS, or have so far remained unreported. Although we do not conduct a formal inter-laboratory comparison either, we do argue that periodic inter-laboratory tests, even on small scales, may be beneficial. In particular, with a larger variety of MALDI-ToF MS instrumentation starting to be used by the ZooMS community, such comparisons might shed light on the benefits of particular mass spectrometry applications for particular ZooMS applications.

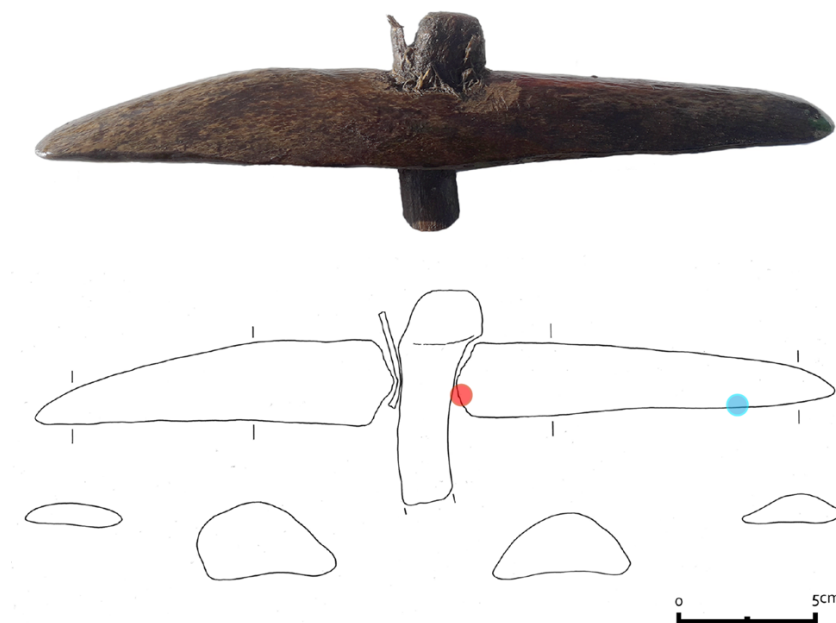


Figure 2 - Photograph of the bone artefact, 2199, and drawing show the cross sections of the tool. The location of the first two samples is indicated by a red circle, the location of the third by a blue circle. © R. Timmermans, Archol, Leiden.

Use-wear methods

Apart from the biomolecular analysis, use-wear analysis was also undertaken in order to investigate the function of the bone tool. Use-wear analysis is based on a visual comparison of the traces on archaeological tools and experimental replicas. The method has been commonly used in archaeology during the last decade (e.g. LeMoine 1994, Van Gijn 2007 and Évora 2015). For the analysis of the object described here a Leica M80 stereomicroscope (magnifications 7.5 - 60 x), a Leica DM2700 metallographic microscope (magnifications 100 and 200 x) and the reference collection of the Leiden University laboratory for Material Culture studies were used. Photos were taken with a Leica MC120HD camera.

ZooMS

Peptides were extracted for ZooMS analysis using both the cold acid and the ammonium bicarbonate (AmBic) extraction protocols (van Doorn et al., 2011). The samples at both York and Leipzig were treated according to similar protocols. In short, the cold acid protocol consists of demineralising the bone fragment in 0.6 M hydrochloric acid (HCl) at 4°C for 18 hours, followed by the removal of the acid supernatant and neutralisation using 50 mM AmBic (NH_4HCO_3 , pH 8.0). The proteins in the samples are

then gelatinised via an incubation in 50 mM AmBic at 65°C for one hour. Subsequently, 50 µL of the resulting supernatant was digested using trypsin (Promega) at 37°C, acidified using trifluoroacetic acid (10% TFA), and then cleaned and desalted on C18 ZipTips (Thermo Scientific). Lastly, the filtered peptides eluted in 0.1% TFA in 50:50 acetonitrile and UHQ water are spotted in triplicate on a MALDI plate along with α -cyano-4-hydroxycinnamic acid (CHCA, Sigma) as a matrix solution. The plate was then analysed using an autoflex LRF MALDI-TOF (Bruker) in Leipzig in reflector mode, positive polarity, and matrix suppression up to 590 Da and collected in the mass-to-charge range 1000–3500 m/z; and a Ultraflex III MALDI-TOF (Bruker) in York in reflector mode, positive polarity, peptide masses below 650 were suppressed and the mass range was set between 800–4000 m/z. Before sample acquisition, external mass calibration was achieved on adjacent MS standard spots with a peptide calibration standard. In Leipzig (#8206195, Bruker Daltonics, Germany) containing a mixture of seven peptides: Angiotensin II m/z = 1046.541, Angiotensin I m/z = 1296.685, Substance_P m/z = 1347.735, Bombesin m/z = 1619.822, ACTH (1–17 clip) m/z = 2093.086, ACTH (18–39 clip) m/z = 2465.198 and Somatostatin m/z = 3147.471. In York a mixture of six peptides: Des-Arg1 Bradykinn m/z = 904.681, Angiotensin I m/z = 1295.685, Glu1-Fibrino-peptide B m/z = 1750.677, ACTH (1–17 clip) m/z = 2093.086, ACTH (18–39 clip) m/z = 2465.198 and ACTH (7–38 clip) m/z = 3657.929.

The AmBic protocol differs from the cold acid protocol by skipping the acid demineralisation step. Instead the bone is immediately incubated in 50 mM AmBic at 65°C for one hour. Afterwards the AmBic protocol is the same as the cold acid protocol in terms of digestion, peptide filtering and mass spectrometry. Consequently, the main difference between the two methods is that the cold acid protocol targets the acid-insoluble collagen fraction of the bone, while the AmBic protocol focuses on the AmBic-soluble component of the collagen.

The obtained mass spectra were matched against the curated database of ZooMS biomarkers from the University of York (Presslee, 2020), as well as specific publications on cetacean biomarkers (Buckley et al., 2009, 2014; Kirby et al., 2013; Speller et al., 2016; Welker et al., 2016). The obtained raw spectra can be found at Zenodo under the identifier 10970629. The observed biomarkers allowed us to assign a family level identification to specimen 2199. In order to further specify the taxonomic identification, it was decided to continue with shotgun-proteomic analysis for this specimen.

LC-MS/MS data acquisition for SPIN data analysis

The “Species by Proteome INvestigation” (SPIN) approach is a recently proposed proteomics workflow leveraging automatic approaches to LC-MS/MS data analysis in association with shorter liquid chromatography separation and DIA or DDA spectral acquisition (Rüther et al., 2022). The workflow employed in this study uses the bioinformatic tools published in SPIN, but follows a different protein extraction protocol. Samples were first processed using the AmBic protocol as outlined above, then demineralized and further processed as the cold acid protocol outlined above, save for a few minor modifications after trypsin digestion. This follows previous strategies employed in ZooMS studies (Welker et al., 2016) and recommendations made for DDA-based SPIN analysis (Mylopotamitaki et al., 2023). After digestion, peptides for each extraction protocol were combined by sample location and loaded onto a single StageTip (Rappsilber et al., 2007) together, prior to mass spectrometry analysis. Peptides were eluted from the StageTips with 30 µL of 40% acetonitrile (ACN) and 0.1% formic acid (FA) and vacuum centrifuged until less than 3 µL remained. The samples were then resuspended in 8 µL of 0.1% trifluoroacetic acid (TFA), 5% ACN. Of each sample 0.5 µL was loaded on an EASY-nLC 1200 (Thermo Fisher Scientific) connected to an Orbitrap Exploris 480 (Thermo Fisher Scientific). Mass spectrometric data was acquired using previously published parameters for archaeological samples (Brandt et al., 2023). A separate blank extraction was performed alongside the adze samples, resulting in a total of three LC-MS/MS injections. The resulting proteomics data was deposited to the ProteomeXchange Consortium via PRIDE (Perez-Riverol et al., 2022) with the dataset identifier PXD051408.

MaxQuant data analysis

Two MaxQuant searches were performed. For the first, the three .raw files were analysed in MaxQuant version 1.6.7.0 against a custom database containing concatenations of the mature COL1A1 and COL1A2 sequences of 54 mammalian species, i.e. the triple helical region and the telopeptides. The database includes 13 cetaceans, of which 8 are protein sequences derived from the genomic resources provided by (Árnason et al., 2018). Non-cetacean sequences were downloaded from NCBI. Variable modifications that were included in the search, were oxidation (M), deamidation (NQ), Gln → pyro-Glu, Glu → pyro-Glu, and proline hydroxylation, with protease specificity set to Trypsin/P in specific cleavage mode.

The second search was performed with MaxQuant version 2.4.10.0 with the same three .raw files, but against a more constrained database (SI 1), consisting of the COL1A1 and COL1A2 sequences of 14 and 12 baleen whale species, respectively. These collagen sequences were acquired from publicly available resources and additional sequences obtained by translating protein sequences for baleen whale genomic resources available online (see Bioinformatics data analysis section below). For this second search oxidation (M), acetyl (protein N-term), deamidation (NQ) and proline hydroxylation were selected as variable modifications. A maximum number of six modifications per peptide was allowed and the protease specificity was set to trypsin/P. A maximum of two missed cleavages was allowed for each peptide. Allowing semi-specific trypsin cleavage can increase the number of detected peptides, especially in more degraded samples where protein fragments are relatively short already before digestion. The enzymatic digestion of these shorter protein fragments creates a relatively large amount of peptides with a semi-specific cleavage pattern, because the other end of the peptide had been broken due to diagenesis rather than enzymatic digestion. Allowing semi-tryptic cleavage in the protein analysis will facilitate the identification of such peptides, but this is not always the case. The increase in search space caused by the inclusion of semi-tryptic peptides may inflate the number of decoy matches, which in turn can result in a reduced number of target peptide matches as a higher score threshold must be used to maintain the same false discovery rate (FDR) (Fahrner et al., 2021; Palomo et al., 2023). We chose to only search for fully tryptic peptides as the initial results indicated relatively good preservation.

Bioinformatics data analysis

Genomic protein translation

Raw reads from 23 baleen whales, representing 15 different species, were accessed from the European Nucleotide Archive or NCBI Sequence Read Archive (SI 2). As genetic data was missing for the Bryde's whale (*Balaenoptera brydei*), we generated a 30x genome from a juvenile male (specimen MCE1246) stranded on September 1st 2000 in Isefjorden, Denmark. DNA was extracted using the Thermo Scientific KingFisher Cell and Tissue DNA Kit and sequenced on a DNBSEQ at BGI China. Accessed reads were trimmed using the bbdut.sh script from bbmap (Bushnell, 2014), with the settings ktrim=r, k=23, mink=8, hdist=1, tbo, qtrim=rl, trimq=15, maq=20, and minlen=40. The blue whale (*Balaenoptera musculus*) reference genome (Bukhman et al., 2024) was selected for read mapping given its status as a platinum-standard reference genome (Morin et al., 2020). Trimmed reads were mapped using the Burrows-Wheeler Aligner Maximal Exact Matches (BWA-MEM) algorithm v.0.7.17 and default settings (Li & Durbin, 2009). PCR duplicates were removed with SAMtools v.1.17 (Li et al., 2009). The target protein transcript sequences were extracted from the mapped (bam) files with the SAMtools view command and genomic coordinates from the reference genome annotations (COL1A1 - CM020960.2: 24504018 - 24520329; COL1A2 - CM020949.2: 58946669 - 58982479). Consensus sequences were generated from the bam files using ANGSD v.0.940-2 (Korneliussen et al., 2014) with the parameters doFasta 2, doCounts 1, minQ 30, and minMapQ 30. To ensure high confidence in the final translated protein sequences, the following steps were manually applied to each consensus sequence individually in the Geneious Prime v.2020.2.5 (Kearse et al., 2012) graphical user interface. The annotated coding sequences for the blue whale COL1A1 (NCBI Accession ID: XM_036835733.1) and COL1A2 (NCBI Accession ID: XM_036862852.1)

genes were each aligned to the corresponding consensus transcript sequences generated for each species with MAFFT v.7.490 (Katoh et al., 2002; Katoh & Standley, 2013). Gaps in the alignment representing untranslated regions (UTRs) were trimmed and the resulting coding sequence from the target species was extracted and translated to the protein sequence in Geneious Prime.

To validate this approach, we compared our resulting protein sequences for those previously annotated for the blue whale reference genome COL1A1 (NCBI Accession ID: XP_036691628.1) and COL1A2 (NCBI Accession ID: XP_036718747.1). Validation also included comparing our protein sequences with the annotated protein sequences from the fin whale (*Balaenoptera physalus*) and North Atlantic right whale (*Eubalaena glacialis*) assembly websites (dnazoo.org/assemblies) generated by the DNazoo consortium (Dudchenko et al., 2017). Through alignment with existing annotated protein sequences for the fin whale and North Atlantic right whale, we observed that the blue whale COL1A1 transcript variant X1 (XP_036691628.1) had an additional 43 amino acids at amino acid positions 433 to 475. This protein sequence region was trimmed from each of our translated protein sequences for consistency with publicly available annotations. Two accessions (SRR13167951; *Eubalaena australis* (southern right whale) and SRR935201; fin whale (*Balaenoptera physalus*), yielded consensus sequences with excessive missing bases and were excluded from the analysis. In addition to the curated sequences, we concatenated 12 publicly available sequences (SI 1) to comprise a final COL1 dataset of 33 baleen whales, representing 14 species.

Mass spectrometric data

Subsequent to spectral identification, data analysis was conducted largely through R (version 4.1.2) in RStudio (version 2022.02.0.0) using the packages tidyverse (version 1.3.1) (Wickham et al., 2019), seqinr (version 4.2-8), ggpubr (version 0.4.0), data.table (version 1.14.2) (Barett et al., 2024), bit64 (version 4.0.5) (Karunaratne et al., 2022), ggsci (version 2.9), progressr (version 0.10.0) (Bengtsson, 2023), stringi (version 1.7.6), MALDIquant (version 1.22.2) (Gibb & Strimmer, 2012), MALDIquantForeign (version 0.14.1) (Gibb, 2024) and gridExtra (version 2.3) (Auguie, 2017). For SPIN data analysis, the peptides detected in the first MaxQuant search (see above) were analysed using the DDA script with default settings and protein sequence database provided by R  ther et al. (2022). This protein database of 20 bone proteins across 156 mammalian species contains 15 cetacean species of which 4 are baleen whale entries. After ZooMS and SPIN, and having determined that the bone artefact probably derives from a baleen whale, we conducted subsequent MaxQuant and amino acid sequence analysis against a sequence alignment of 29 baleen whale sequences, representing 14 species for COL1A1 and 25 sequences, representing 12 species for COL1A2. An overview of which whale genomes provided reference for COL1A1 and/or COL1A2 can be found in SI 2. We determined all relevant single amino acid polymorphisms (SAPs) at which species of the baleen whale genera *Balaena* and *Eubalaena* differ, and compared these SAPs with the coverage obtained in the second MaxQuant database search.

Finally, the deamidation rate of the proteins observed during the targeted MaxQuant search, excluding any contaminants, was calculated using the deamidation tool (Mackie et al., 2018).

Results

Use-wear results

Inspection of the bone artefact shows that the tool was formed from a longitudinal section of the bone, as the trabecular bone is visible. According to the observed use-wear traces, the first shaping was done by cutting, chopping and grinding (Figure 3A). Exactly how the rounded ends were shaped is unclear, as any traces of this were removed by subsequent grinding to give them their final shape. The hole is square in shape and was made using a chisel (Figure 3B), working from both sides of the tool. On the inside of the hole, clear wear of a shaft is visible, indicating prolonged use. The tool was bound to the

haft using a rope that extended to about midway between the hole and the edges of the tool. The traces left by the binding materials show both characteristics of plant and animal materials (Figure 3C). This can either indicate the use of bast as a binding material or the use of different binding materials in different stages of the tools use-life. This could not be interpreted with certainty. The traces on the edges of the tool are unfortunately less clear and one of the edges is damaged, which removed most of the use-wear traces. The traces that are visible are fatty, and in some spots smooth in character. The edge is rounded and there are clear striations visible in the polish (Figure 3D). This leads to the interpretation that the tool was used on a plant material. Based on the rounding of the edge it was not used to chop, but more probably the pounding or the processing of fibres in another manner.

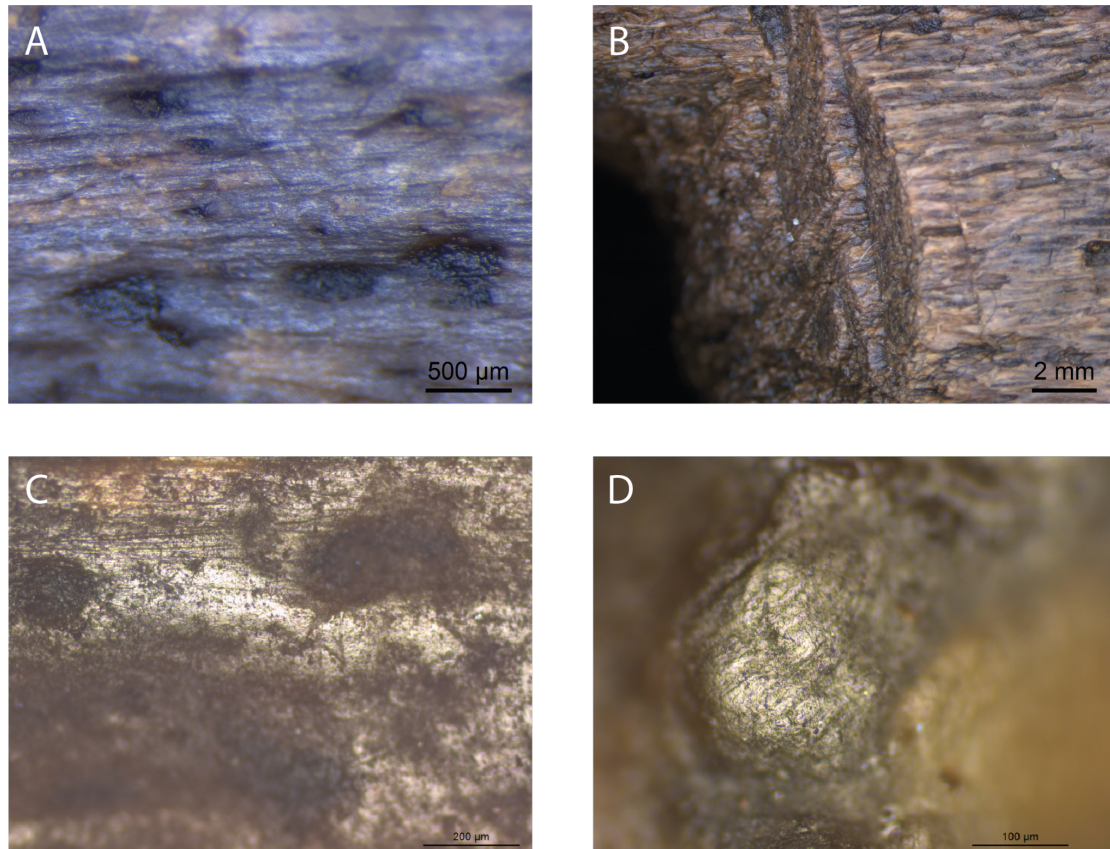


Figure 3 - Photos of the use-wear on the bone object. A, traces of grinding on the lower part of the tool (original magnification 40x). B, traces of chiselling in the hole (original magnification 7,5x). C, traces of the binding material used to bind the tool to the haft (original magnification 100x). D, traces on the edge of the tool (original magnification 200x).

ZooMS

ZooMS extraction was performed independently in two different locations and analysed on separate MALDI TOF-MS instruments. Three out of four extractions of the bone artefact provided consistent results attributing the sample to *Eubalaena glacialis* (North Atlantic right whale) or *Balaena mysticetus* (bowhead whale) (Table 1). The fourth sample provided inconsistent results with marker masses observed that suggest a mixture of taxonomic sources, preventing an adequate taxonomic identification through ZooMS for this extraction. Nevertheless, the consistency of the marker masses observed among the other three extracts (Figure 4), generated in two laboratories and analysed on different MALDI TOF-MS instruments, gives confidence that according to ZooMS the bone artefact likely belongs to the Balaenidae family.

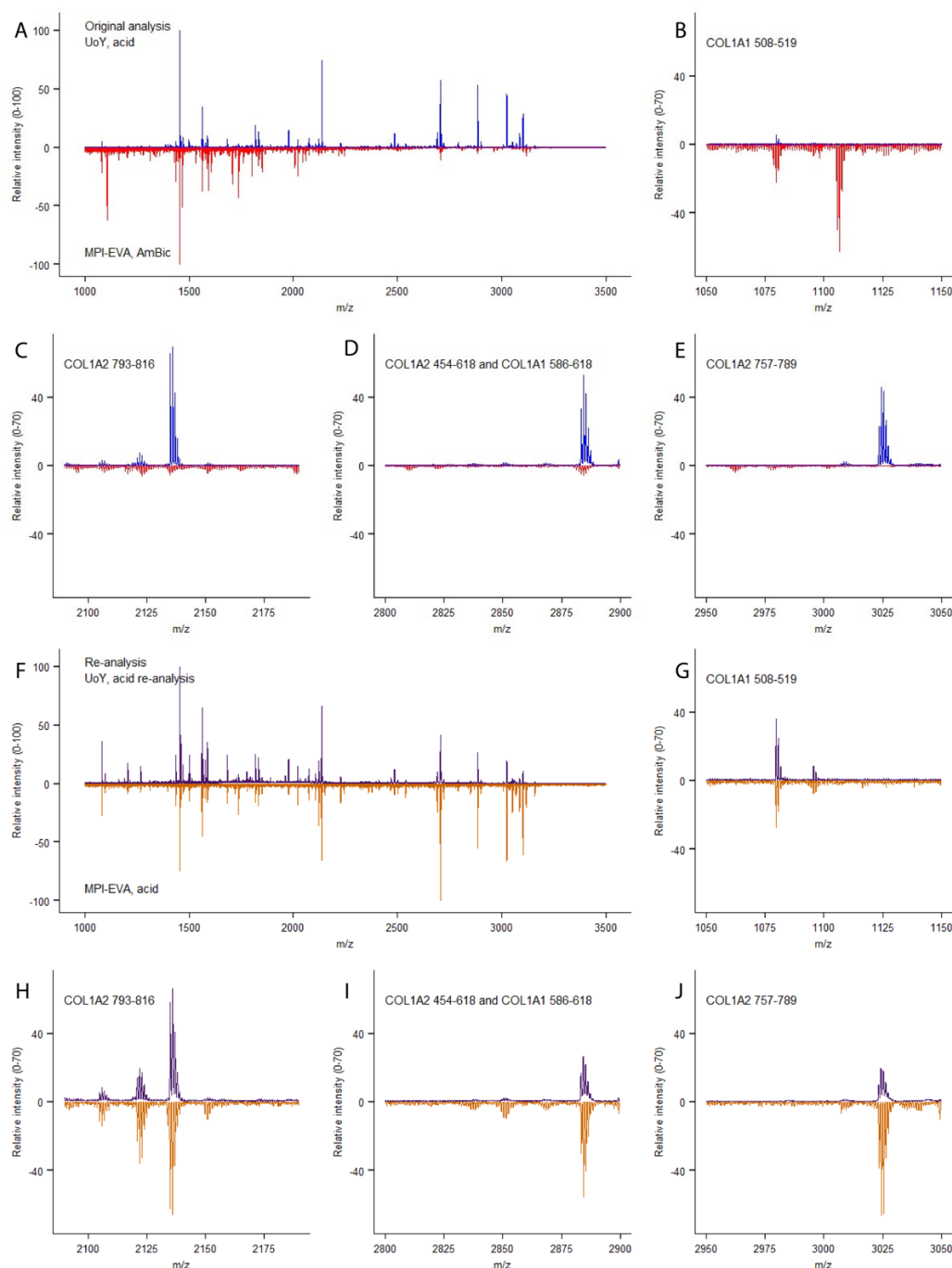


Figure 4 - ZooMS spectra. A, total spectra of the original UoY (blue) and MPI-EVA (red) analyses. B-E, close-ups of the mass spectra for biomarkers COL1A1 508-519, COL1A2 793-816, COL1A2 454-618 AND COL1A1 586-618, COL1A2 757-789 respectively. F, total spectra of the re-analysed samples from UoY (purple) and MPI-EVA (orange). G-J, zoomed in spectra of the biomarkers COL1A1 508-519, COL1A2 793-816, COL1A2 454-618 AND COL1A1 586-618, COL1A2 757-789 respectively.

Table 1 - Observed biomarkers and matching taxa, oxidised variants of biomarkers are notated in parentheses, * indicates that a possible peak at 1161.6 m/z was also observed, although at a lower intensity than 1189.6 m/z. UoY denotes a protein extraction performed at the University of York, whereas MPI-EVA indicates samples extracted at the Max Planck Institute, Department for Evolutionary Anthropology, Leipzig. “Acid” signifies that the sample was extracted using the cold acid extraction (van Doorn et al., 2011), “AmBic” indicates an AmBic buffer extraction protocol (van Doorn et al., 2011). The NA right whale refers to the North Atlantic right whale (*Eubalaena glacialis*).

Sample	COL1a1 508 - 519	COL1a2 978 - 990 / (+16)	COL1a2 484 - 498	COL1a2 502 - 519	COL1a2 292 - 309	COL1a2 793 - 816	COL1a2 454 - 483	COL1a1 586 - 618 / (+16)	COL1a2 757 - 789 / (+16)	Latin name	Common name
UoY, acid, 2199	1079.6	1205.6	1453.7	1566.8	1682.8	2135.1		2883.4/ 2899.4	3007.4 / 3023.4	<i>Eubalaena glacialis</i> / <i>Balaena mysticetus</i>	NA right whale / Bowhead whale
UoY, acid, reanalysis, 2199	1079.6	1189.6/ 1205.6	1453.7	1566.8	1682.8	2135.1	2849.4	2883.4 / 2899.4	3007.4 / 3023.4	<i>Eubalaena glacialis</i> / <i>Balaena mysticetus</i>	NA right whale / Bowhead whale
MPI-EVA, AmBic, 2199	1079.6/ 1105.6	1161.6	14537	1550.7/ 1566.7	1648.8/ 1682.8	2115.8	2808.3	2883.4 / 2899.4	2999.2	Contaminated	-
MPI-EVA, acid, 2199	1079.6	1189.6*/ 1205.6	1453.7	1566.8	1682.8	2135.1	2849.4	2883.4 / 2899.4	3007.4 / 3023.4	<i>Eubalaena glacialis</i> / <i>Balaena mysticetus</i>	NA right whale / Bowhead whale

SPIN data analysis

Since the ZooMS taxonomic assignment of the artefact specimen 2199 leaves several baleen whale species as the possible taxonomic origin, we proceeded with separate proteomic extractions for shotgun proteomic analysis. One extraction included material sampled from the inside of the hafting hole of the artefact, while the other extraction included material sampled from the outside of the artefact. The generated data was first analysed using the SPIN analysis workflow and the SPIN mammalian bone protein database, which managed to identify the amino acid sequences for 419 peptides from the outside of the bone tool and 469 peptides from the inside of the artefact. These peptides were matched to a variety of different proteins with varying degrees of taxonomic specificity. Taxonomic identifications (Table 2) were obtained by evaluating the frequencies of proteins matching a particular taxon, resulting in a SPIN assignment to *Balaenoptera acutorostrata*. The extraction blank was matched to *Bos* sp. by the SPIN script, but it has been excluded from Table 2 as 84% of the peptide spectral matches (PSMs) in the blank were derived from the trypsin used for digestion. This also means that the number of PSMs used by the SPIN script to assign a taxonomic identification to the blank sample is very low and most likely derived from background contamination. The blank intensity signal can therefore be regarded as background noise.

Table 2 - SPIN taxonomic identifications of the bone tool (2199). The Site count refers to the number of amino acid positions called with high confidence and present in the protein sequences of the highest-ranking species entry (here, *Balaenoptera acutorostrata*). The Non-matching site count indicates the remaining high-confidence amino acid positions that do not match the highest-ranking species entry.

Sample	Latin name	Common name	Site count	Non-matching site count	Relative protease intensity (%)
2199, inside hafting hole	<i>Balaenoptera acutorostrata</i>	Common minke whale	2310	159	1.1
2199, lower side artefact	<i>Balaenoptera acutorostrata</i>	Common minke whale	2392	169	1.0

At first glance it seems that the SPIN results, an assignment to *Balaenoptera acutorostrata*, contradict the ZooMS results, with an assignment to either *Eubalaena glacialis* or *Balaena mysticetus*. However, the SPIN database only contains entries for four closely-related baleen whale species within the genus *Balaenoptera*, whereas the ZooMS peptide marker database includes entries for eight baleen whale species, including *Eubalaena glacialis*, *Balaena mysticetus*, and *Balaenoptera acutorostrata*. This suggests the SPIN taxonomic identification might be driven by a lack of representative sequence entries in case the “true” taxonomic identity lies outside the four species included within its database.

Selected baleen whale database search

To resolve this potential conflicting taxonomic identification, we translated COL1A1 and COL1A2 sequences from genomic resources available for 14 baleen whale species. Next we performed a second MaxQuant search against the resulting database of all available eight COL1A1 and five COL1A2 baleen whale sequences plus the additional sequences translated from the acquired genomic data for a total of 14 and 12 species for COL1A1 and COL1A2, respectively. Our COL1A1 and COL1A2 sequence database therefore contained a sequence entry for each known baleen whale genus and species, except *Balaenoptera omurai* (Omura’s whale) and *Balaenoptera edeni* (Eden’s whale, a small form of Bryde’s whale).

We reanalysed our shotgun proteomic data against this new baleen whale-specific sequence database containing COL1 entries only. The protein groups identified in this more targeted search were filtered by removing any matches to a decoy sequence, as well as those protein groups that only had two

or fewer unique peptides. Additionally, we monitored the prevalence of deamidation using a publicly available Python script (Mackie et al., 2018) as a rough indicator of modern contamination. The results indicate that on average 52.7% of the asparagine residues in the bone object were deamidated, as well as 28.4% of the glutamine residues. These values fall in the overlapping range of previously published deamidation values for modern and archaeological samples (Ramsøe et al., 2020; Pal Chowdhury & Buckley, 2022). However, the standard deviation of deamidation in modern bones is so large that even a sample from a 46-107 ka cave site fell within their range (Brown et al., 2021). The deamidation values therefore raise no doubts regarding the age of the extracted collagen.

These criteria left one protein group for each of the two proteins included in the database (Table 3). Both the protein groups contain collagen variants of three different species. The COL1A1 protein group contains peptides matching to the three species of the genus *Eubalaena*, while the COL1A2 protein group also includes the species *Balaena mysticetus*. Aligning the COL1A1 and COL1A2 sequences of *Eubalaena* sp. and *Balaena mysticetus* reveals five SAPs that would allow distinguishing between the two taxonomic groups. For each of the five SAPs the number of PSMs matching the *Eubalaena* sp. and the *Balaena mysticetus* reference sequence were counted to provide an overview of the support for both taxa (Table 4). The taxonomic specificity of the five SAPs was checked by searching the longest observed peptide covering each SAP against the NCBI nr database using BLAST (Altschul et al., 1990). Though it must be noted that these five peptides were not selected for being unique to either *Eubalaena glacialis* or *Balaena mysticetus*, but for differing between the two species. The ZooMS and SPIN-analysis results enable us to refine the taxonomic identification starting from the level of baleen whales. Nevertheless, it is important to check the specificity of the target peptides to see if they are shared with common contaminant taxa.

Table 3 - Number of peptides per species for the most abundant protein group for COL1A1 and COL1A2.

Protein	Species	Peptide count (razor + unique)
COL1A1	Eubalaena australis	135
COL1A1	Eubalaena japonica	135
COL1A1	Eubalaena glacialis	135
COL1A2	Eubalaena glacialis	109
COL1A2	Eubalaena japonica	109
COL1A2	Balaena mysticetus	109

Table 4 - SAPs differing between *Eubalaena* sp. and *Balaena mysticetus*, and the number of matching observed PSMs.

Protein	Species	Position	AA before	SAP	AA after	N PSMs
COL1A1	Balaena mysticetus	935	G	S	P	11
COL1A1	Eubalaena sp.	935	G	A	P	48
COL1A1	Balaena mysticetus	1193	S	A	G	0
COL1A1	Eubalaena sp.	1193	S	V	G	46
COL1A2	Balaena mysticetus	141	K	G	G	0
COL1A2	Eubalaena sp.	141	K	A	G	7
COL1A2	Balaena mysticetus	276	P	S	G	0
COL1A2	Eubalaena sp.	276	P	P	G	10
COL1A2	Balaena mysticetus	935	G	A	P	0
COL1A2	Eubalaena sp.	935	G	P	P	25





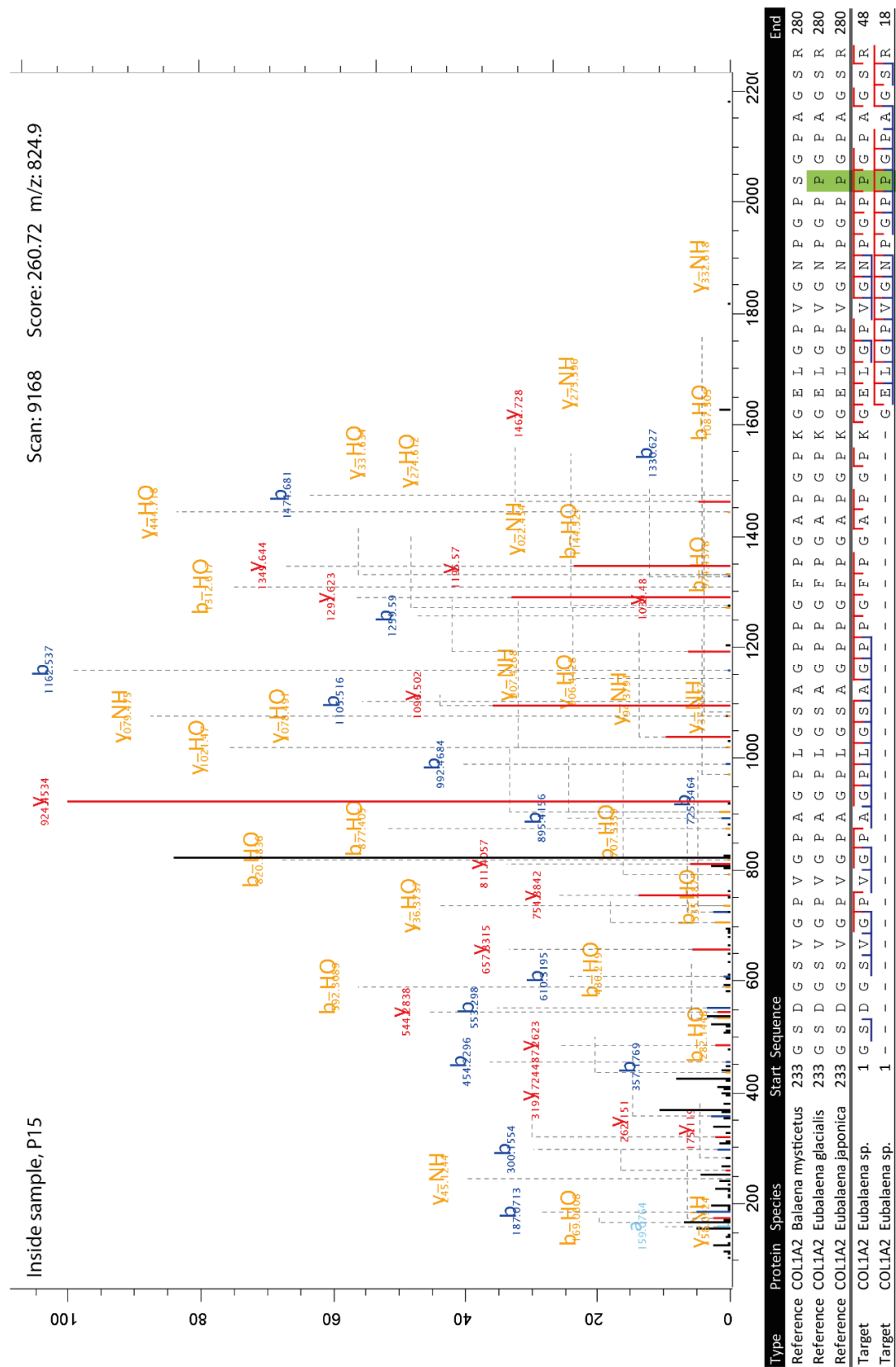


Figure 7 - Highest scoring MS2 spectrum of the COL1A2 262-280 peptide covering the 276 SAP. The reference spectra matching the observed peptides are highlighted in green at the SAP. The b-ion coverage is annotated in blue, and the y-ion coverage in red.

Of the five, the COL1A1 peptides covering the SAP at position 935 and 1193 did not match to any other species. Unfortunately, BLAST obtained no significant match for the COL1A2 peptide covering the SAP at position 141 and the peptide covering the SAP at position 276 was shared with *Balaenoptera acutorostrata*, *Balaenoptera ricei* and *Balaenoptera musculus*. Lastly, the COL1A2 peptide covering the SAP at position 935 was found to be shared among several other taxonomic groups. These were several species of bats (Chiroptera), three Metatherian species and 13 toothed whale species (Odontoceti). However, the only baleen whale species it matched was *Eubalaena glacialis*. It should be emphasised that although BLAST is often considered the golden standard for taxonomic identification, the NCBI nr database it relies on also suffers from incomplete references. Consequently, a custom database, as used here, may be able to achieve more precise taxonomic identifications.

For four of the five SAPs the observed peptides only matched with the *Eubalaena* sp. sequences, but for the COL1A1 SAP at 935 there were also a number of PSMs matching the *Balaena mysticetus* sequence. To further validate the peptide identifications, the MS2 spectra and alignment of the respective peptides of the SAPs at position 935, 1193 for COL1A1 and at position 276 for COL1A2 were visualised (Figures 5-7). These show that the fragment ion coverage of the SAPs is quite extensive for all the relevant peptides and argues in favour of the validity of these peptide sequence identifications, including the contrasting sequences for the SAP at COL1A1 935. The majority of the protein evidence suggests *Eubalaena* sp. as the source of the bone artefact (Specimen 2199), as for all SAPs several peptides matching to the *Eubalaena* sp. reference were found.

Discussion

Contextualising molecular taxonomic identification

Three proteomic workflows have been applied in this study, each with different levels of specificity and requiring different approaches to critically interpret their results (Figure 8). A common feature is that in the initial phase of analysis no a priori assumptions on the presence or absence of species are made, but all species are considered as possibilities. The taxonomic identification should be made on a taxonomic level encompassing all species matching the proteomic signal, but excluding lower-level taxa with their own reference data that contradict the observed data. The baleen whale clade illustrates this principle well, as no reference data is available for several of its members, for both ZooMS and SPIN. Consequently, for example in the case of the ZooMS analysis of sample 2199, *Eubalaena australis* (southern right whale) could also be considered as a candidate, because no collagen biomarkers have been established for this species. It could be that the collagen sequence that sets *Eubalaena australis* apart from the *Eubalaena glacialis* and *Eubalaena japonica* sequences was ancestral to all members of *Eubalaena*. As a result, the other members of the *Eubalaena* genus for which we do not have reference biomarkers should be taken into consideration. Thus, to minimise the impact of limited reference data in proteomic identification of bone specimens, we constructed a database of collagen type I sequences capturing most of the existing sequence variation among extant baleen whales for the targeted database search. This database, which misses only *Balaenoptera edeni* and *Balaenoptera omurai*, could in turn be used to significantly expand the ZooMS peptide marker database. Additionally, the taxonomic resolution of the proteomic identifications can be greatly improved by filtering them based on the archaeological and geographical context of the sample.

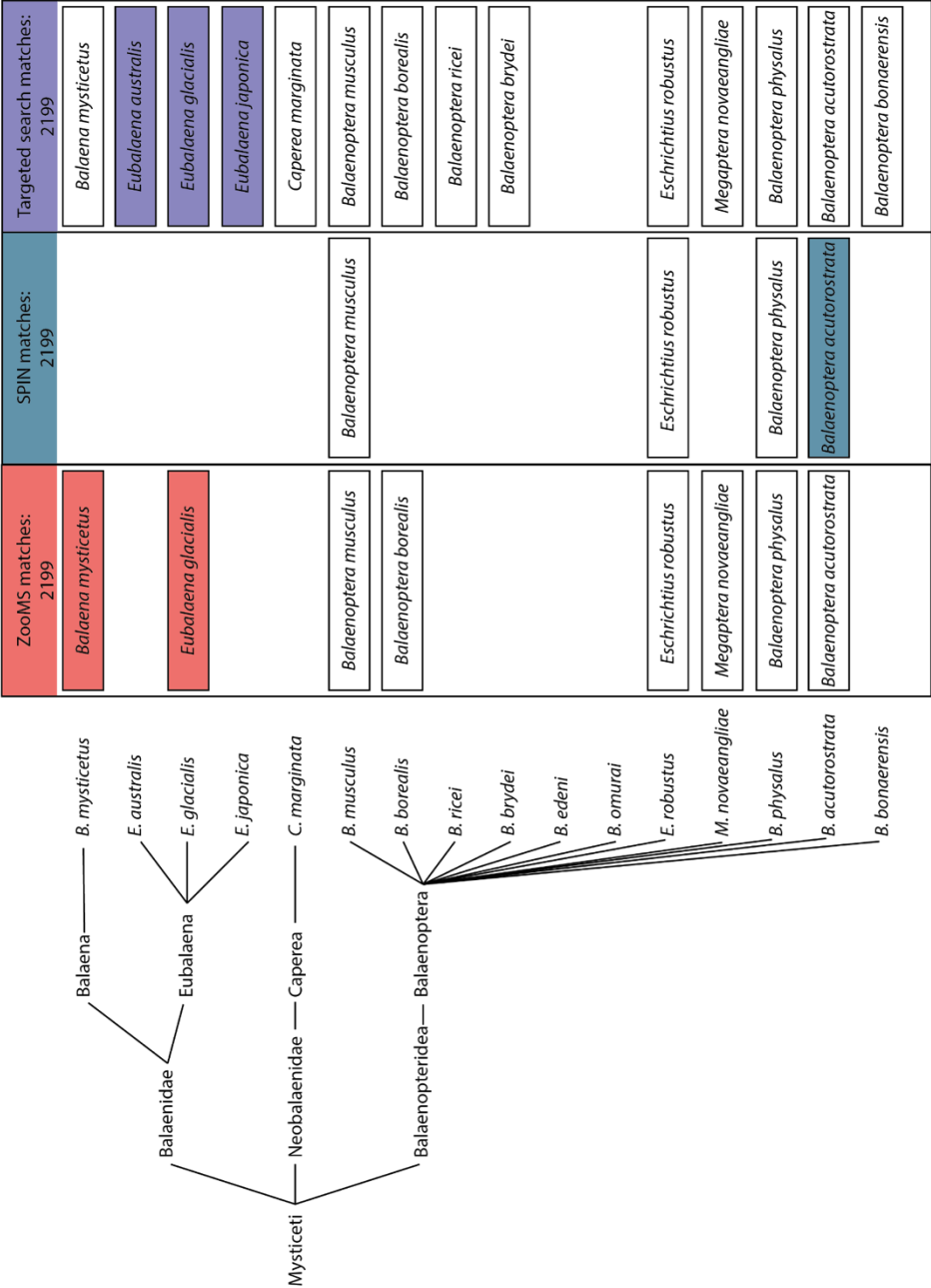


Figure 8 - Baleen whale phylogeny and presence in relevant proteomic databases. Left: the phylogenetic tree of baleen whales based on recent genomic studies (Árnason et al., 2018; McGowen et al., 2020; Wolf et al., 2023). Right: tabular overview of the species that can be taxonomically identified using ZooMS and SPIN. Species for which reference biomarkers are available have a black border and those that match the observed biomarkers following their analytical workflow have a coloured background. Species absent for ZooMS, SPIN, and/or the targeted search are left out from each relevant column.

Here, we assess potential database biases and review the historical distributions of the possible species identified by the molecular analysis to exclude species that were either identified as false positives due to database limitations or not present in the North Sea during the Holocene (Figure 8). First of all, while the species is present in the North Sea, we argue that the *Balaenoptera acutorostrata* identification for the bone artefact specimen 2199 obtained by SPIN can be excluded from the possible identifications, because it is an artefact of the database selection. No Balaenidae were included in the SPIN database, whereas all except two baleen whale species were included in the database used in the targeted search. As both the ZooMS and targeted search analyses of specimen 2199 used more comprehensive databases, including Balaenidae and Balaenopteridae references, the conflicting *Balaenoptera acutorostrata* identification can be safely regarded as a database artefact.

As for the remaining list of potential species candidates for specimen 2199, the archaeological context of Heiloo Zuiderloo allows the taxonomic identification to be narrowed down to a single species; *Eubalaena glacialis* (North Atlantic right whale). Ideally this identification would be based on securely identified remains from the same region and time period. However, due to the limited availability of species-level identified whales from Bronze Age Netherlands, we included archaeological findings from other periods, as well as the modern distribution of the candidate species. The presence of *Eubalaena glacialis* is attested in the Netherlands from the mid-Holocene (Foote et al., 2013) and at least the first millennium AD onwards (van den Hurk et al., 2022) and it has been tentatively suggested that this species might have been one of the most frequently exploited cetaceans along the European Atlantic coast (van den Hurk et al., 2023). The two other candidate species, *Eubalaena japonica* (North Pacific right whale) and *Eubalaena australis* (southern right whale), are native to the North Pacific Ocean (Kenney, 2009) and the southern hemisphere, respectively (Richards, 2009) and can therefore be safely excluded as potential sources of sample 2199. Instead, it seems most likely that the adze was produced from a bone of *Eubalaena glacialis*. The proteomic taxonomic identification of the two objects deviates substantially from the original morphological estimation. However, it seems that for fragmented remains it can be particularly difficult to distinguish Elephantidae and Mysticeti. In fact, several cases of biomolecular techniques unmasking morphologically identified Elephantidae remains are known (van den Hurk et al., 2020), highlighting that this may be an area where accessible biomolecular taxonomic identifications, such as ZooMS, are particularly useful.

Bronze Age whale exploitation: hunting or scavenging?

Prehistoric baleen whale remains were until recently considered rare in the North Sea region, but are increasingly reported. Aaris-Sørensen et al. (Aaris-Sørensen et al., 2010) identify at least 100 baleen whale Pleistocene and Holocene remains in Denmark with several dating to the Bronze Age, and Foote et al. (Foote et al., 2013) documented the presence of North Atlantic right whales and bowhead whales in the Netherlands back to the early Holocene and Pleistocene, respectively. These finds make it clear that the North Sea region has a long history of baleen whale exploitation, but whether this consisted of the opportunistic use of beached whales or active hunting remains a question for a large part of this history. It is impossible to determine if a single bone is derived from a beached or hunted animal, but several characteristics have been proposed that would suggest it is more likely that a whale bone assemblage was obtained by active whaling rather than opportunistic scavenging. Active whaling assemblages are thought to be relatively limited in taxonomic diversity (Wellman et al., 2017; van den Hurk et al., 2023), whale remains must be abundant and there must be some form of industry associated with the whale remains, such as bone artefact production or blubber processing (Wellman et al., 2017; Hennius et al., 2018, 2023). A potential side-effect of this requirement is that active whaling purely for consumption becomes more difficult to detect archaeologically than whaling to acquire resources for product manufacture. Lastly, written sources are frequently used as evidence for active whaling in historical periods (Hennius et al., 2018; van den Hurk et al., 2023), but these are not available for

prehistoric periods. In general, most studies on active whaling agree that it needs to be accompanied by a certain scale of whale resource exploitation.

Early suggested examples of active whaling are the Dutch Neolithic Vlaardingen Culture (van den Hurk et al., 2023) and 6th century Scandinavia (Hennius et al., 2018). Then from the Middle Ages onwards active whaling seems to become more prevalent along the Atlantic coast, most notable are the Basques and northern Spaniards starting from at least the 11th century (Rey-Iglesia et al., 2018), but also in the Netherlands (van den Hurk et al., 2022). The evidence for active whaling in the Vlaardingen Culture consists of 13 and 28 baleen whale bones at two sites respectively, 10 and 7 of which were identified as *Eschrichtius robustus* (the grey whale). The dominance of *Eschrichtius robustus* suggests a taxonomic focus characteristic of active whaling. Additionally, the sites indicate an overall marine focus of the culture, displayed by their diet and material culture (Brinkkemper et al., 2011; van den Hurk et al., 2023). The case for active whaling in 6th century Scandinavia is more robust. The large increase in the production of whale bone gaming pieces across Scandinavia, as well as an increase in blubber processing pits indicates a significant increase in the exploitation of whale resources (Hennius et al., 2018). Additionally, most of the whale bone species were assigned to *Balaenidae* sp., again suggesting a hunting preference for a particular taxon (Hennius et al., 2023). For the mediaeval active whaling cultures of the Basques and Northern Spaniards and the Dutch there are also historical sources attesting to active whaling next to the archaeological bone assemblage. Both these cultures also seem to have favoured *Eubalaena glacialis* as their target for whaling (Rey-Iglesia et al., 2018; van den Hurk et al., 2022).

The aforementioned examples of active whaling can be used to evaluate the baleen whale presence in the archaeological record of the Dutch Bronze Age. Apart from the *Eubalaena glacialis* artefact described in this study, eight other cetacean remains have been found at the same site of Heiloo Zuiderloo (van der Heiden, 2018; de Koning & Tuinman, 2019; Moesker et al., 2021), one of which was also identified as *Balaenidae* and two others as grey whales (van den Hurk et al., 2023). Additionally, several Bronze Age whale remains have been found relatively close to Heiloo. Two whale ribs dated to 1550-1350 BC have been found near Bergen (de Ridder, 1993), another whale rib was found in Boekelermeer, dated to around 1500 BC, and two whale bones were found in Alkmaar, one being dated to the Middle Bronze Age (Kleijne, 2015). The ribs from Bergen and Boekelermeer have recently been identified using ZooMS as sperm whale (*Physeter macrocephalus*) and grey whale (*Eschrichtius robustus*) respectively (van den Hurk et al., 2023). The remains from Alkmaar could not be identified beyond 'whale', which likely is supposed to refer to any large-bodied cetacean. Regarding whether these cetacean remains are the result of active whaling or scavenging, in the case of the Bergen ribs they were said to have been found in a marine deposit. The authors hypothesise that these two ribs are the remains of a stranded whale butchered on the beach where it landed (de Ridder, 1993). Considering the remaining finds, the evidence for active whaling does not appear strong. Although there is a cluster of cetaceans from a relatively narrow time period in a small area, they feature a high taxonomic diversity more indicative of opportunistic exploitation of beached whales. We cannot exclude active whaling, but neither does there appear to be strong evidence for the presence of active whaling in the Dutch Bronze Age.

Contextualising the tool in the Bronze Age

The current study has mainly been devoted to studying the first step in a bone tool biography: the choice of the raw material for the tool. However, the find location of the tool raises questions regarding the last chapter in its biography: its deposition, particularly the intentionality behind the deposition. The object was found near a local depression in the dune landscape. Extensive studies have shown that in Bronze Age Europe, these natural wet locations were desirable places for intentional deposition and 'destruction' of what we consider valuable objects (Fontijn, 2002, 2020). In the vicinity of the 2020 excavation location, objects associated with such a phenomenon have been discovered in the past.

These include a group of four Late Bronze Age flint sickles and a bronze sickle together inserted upside down in the peat (Brunsting, 1962) and an unused Middle Bronze Age stop-ridge axe that was probably made in Normandy (Moesker et al., 2021). Although the Heiloo specimen was made of an ‘exotic’ material, use-wear traces show that it has been extensively worked and used. Efforts have been made to shape and finish the object and long-term use has left traces on both the bone part and the wooden handle of the object. Although the artefact is not broken and does not appear to have been irreparably damaged, it ultimately ended up in an active watering hole that was emptied at a later time. The evidence suggests that it is not a deliberate deposition, but rather a lost object, although the first cannot be ruled out.

The adze-like tool may be lost in more ways than one. No parallel of the object is currently known. Tools made of bone or antler with a distinct working edge are known from many locations in the Netherlands during the Bronze Age (Zeiler et al. forthcoming). Although no clear parallel for this object can be found, it is notable that in recent years, several other utensils associated with the processing of organic materials such as flax and fibres have been found in Heiloo (Edmonds, 2018; Verbaas & Edmonds, 2018; Koning & Tuinman, 2021; Lange, 2021; Tuinman et al., 2022). These involve tools made of organic materials, including a bone comb and two oak wood pestles. It demonstrates the significant importance of the prehistoric sites in Zuiderloo and the unique preservation conditions in which such organic tools are preserved, but at the moment it does not provide much clarity regarding the intended purpose of the tool nor its place within the Bronze Age toolkit.

Conclusion

Although in recent years advances in biomolecular applications to archaeology have facilitated the taxonomic identification of osseous tools, there remain some taxonomic and methodological challenges. This small scale study on an exceptional bone artefact from Bronze Age Heiloo exemplifies well how biomolecular methods can improve our understanding of what species were exploited, as well as the different challenges and pitfalls that it entails. After performing ZooMS, SPIN, and shotgun-proteomic data analysis and while considering their individual methodological strengths and limitations, we have shown that the extraordinary adze-like bone tool was most likely manufactured from the bone of *Eubalaena glacialis* (North Atlantic right whale). Use-wear analysis furthermore demonstrates that it was used for the processing of plant fibres. We cannot be certain whether the bone to produce the artefact was obtained from a beached whale or by active whaling, but our current understanding of cetacean remains in the Dutch Bronze Age does not point towards the existence of an active whaling practice. Additionally, we contribute previously unavailable COL1A1 and COL1A2 reference sequences for six baleen whale species, in order to facilitate the future proteomic identification of these species within the archaeological and palaeontological record.

Acknowledgements

We would like to thank André Ramcharan and Dr. Laura Llorente Rodriguez for their initial morphological identification and Prof. Marie Soressi for her advice on setting up this project. Special thanks go to Femke Reidsma, Erica van Hees, Dr. Rachel Schats and Dr. Jason Laffoon for their aid in sampling the adze. We thank the IZI Fraunhofer, S. Kalkhof and J. Schmidt for providing access to the MALDI-TOF MS instrument in Leipzig, Germany and the York Centre of Excellence in Mass Spectrometry for the use of the Ultraflex III MALDI-ToF/ToF instrument. Lastly, we would also like to express our gratitude to Meaghan Mackie for her work in acquiring the LC-MS/MS data. Preprint version 2 of this article has been peer-reviewed and recommended by Peer Community In Archaeology (<https://doi.org/10.24072/pci.archaeo.100501>; Goude, 2024).

Funding

J.D. is funded by a Marie Skłodowska-Curie grant (grant agreement No. 956351) from the EU Horizon 2020 programme.

This research has been made possible through funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement no. 948365 (PROSPER) awarded to F.W.

D.M. received support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 861389 (PUSHH).

M.L.M. is funded by a Carlsberg Foundation Semper Ardens Accelerate Grant CF21-0425, awarded to M.T.O.

V.S.-M. is supported by a Fyssen Foundation postdoctoral fellowship (2023-2025). ZooMS analyses performed in Leipzig were financially supported by the Max Planck Society.

The excavation at Heiloo-Zuiderloo and subsequent analysis by J.B. were performed as part of the commercial activities of Archol BV commissioned by the municipality of Heiloo. The use-wear analysis by A.V. in turn was commissioned by Archol BV.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, scripts, code, and supplementary information availability

The MALDI-MS spectra used for ZooMS can be found on Zenodo with the following identifier: <https://doi.org/10.5281/zenodo.10970628> (Dekker, 2024).

The LC-MS/MS data was deposited on the ProteomeXchange Consortium via the PRIDE partner repository under the dataset identifier PXD051408.

References

- Aaris-Sørensen K, Rasmussen KL, Kinze C, Petersen KS (2010) Late Pleistocene and Holocene whale remains (Cetacea) from Denmark and adjacent countries: Species, distribution, chronology, and trace element concentrations. *Marine Mammal Science*, 26, 253–281. <https://doi.org/10.1111/j.1748-7692.2009.00356.x>
- Adamczak K, Kowalski Ł, Urbanová P, Douka K, Brown S, Kowalski AP, Kukawka S, Wilczyński J, Szczepańska G, Zawilski P (2021) New evidence for deer valorisation by the TRB farmers from Poland using ZooMS and micro-CT scanning. *Journal of Archaeological Science: Reports*, 40, 103230. <https://doi.org/10.1016/j.jasrep.2021.103230>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology*, 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Árnason Ú, Lammers F, Kumar V, Nilsson MA, Janke A (2018) Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science advances*, 4, eaap9873. <https://doi.org/10.1126/sciadv.aap9873>
- Augue B (2017) gridExtra: Miscellaneous Functions for “Grid” Graphics. <https://doi.org/10.32614/CRAN.package.gridExtra>
- Barett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, Hocking T (2024) data.table: Extension of ‘data.frame’. <https://doi.org/10.32614/CRAN.package.data.table>

- Bengtsson H (2023) progressr: An Inclusive, Unifying API for Progress Updates. <https://doi.org/10.32614/CRAN.package.progressr>
- Bradfield J, Forssman T, Spindler L, Antonites AR (2019) Identifying the animal species used to manufacture bone arrowheads in South Africa. *Archaeological and anthropological sciences*, 11, 2419–2434. <https://doi.org/10.1007/s12520-018-0688-5>
- Bradfield J, Kitchener AC, Buckley M (2021) Selection preferences for animal species used in bone-tool-manufacturing strategies in KwaZulu-Natal, South Africa. *PloS ONE*, 16, e0249296. <https://doi.org/10.1371/journal.pone.0249296>
- Brandt LØ, Mackie M, Daragan M, Collins MJ, Gleba M (2023) Human and animal skin identified by palaeoproteomics in Scythian leather objects from Ukraine. *PloS ONE*, 18, e0294129. <https://doi.org/10.1371/journal.pone.0294129>
- Brattinga JJ (2023) Opgraving Heiloo-Zuiderloo; Deelplan 3 & Middenduin Bewoningsresten uit de bronstijd en sporen uit de Romeinse tijd tot en met de Nieuwe tijd op de oostelijke flank van de strandwal in Heiloo, gemeente Heiloo.
- Brinkkemper O, Drenth E, Zeiler JT (2011) An outline of the subsistence of the Vlaardingen culture from the Netherlands. *La Revue medicale de Picardie*, 28, 207–220. <https://doi.org/10.3406/pica.2011.3331>
- Brown S, Kozlikin M, Shunkov M, Derevianko A, Higham T, Douka K, Richter KK (2021) Examining collagen preservation through glutamine deamidation at Denisova Cave. *Journal of archaeological science*, 133, 105454. <https://doi.org/10.1016/j.jas.2021.105454>
- Brunsting H (1962) De sikkels van Heiloo. *Oudheidkundige mededelingen uit het Rijksmuseum van Oudheden te Leiden*, XLIII, 107–112.
- Buckley M, Collins M, Thomas-Oates J, Wilson JC (2009) Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 23, 3843–3854. <https://doi.org/10.1002/rcm.4316>
- Buckley M, Fraser S, Herman J, Melton ND, Mulville J, Pálisdóttir AH (2014) Species identification of archaeological marine mammals using collagen fingerprinting. *Journal of archaeological science*, 41, 631–641. <https://doi.org/10.1016/j.jas.2013.08.021>
- Bukhman YV, Morin PA, Meyer S, Chu L-F, Jacobsen JK, Antosiewicz-Bourget J, Mamott D, Gonzales M, Argus C, Bolin J, Berres ME, Fedrigo O, Steill J, Swanson SA, Jiang P, Rhie A, Formenti G, Phillippy AM, Harris RS, Wood JMD, Howe K, Kirilenko BM, Munegowda C, Hiller M, Jain A, Kihara D, Johnston JS, Ionkov A, Raja K, Toh H, Lang A, Wolf M, Jarvis ED, Thomson JA, Chaisson MJP, Stewart R (2024) A High-Quality Blue Whale Genome, Segmental Duplications, and Historical Demography. *Molecular Biology and Evolution*, 41, msae036. <https://doi.org/10.1093/molbev/msae036>
- Bushnell B (2014) BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).
- Cleland TP (2018) Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *Journal of proteome research*, 17, 3976–3983. <https://doi.org/10.1021/acs.jproteome.8b00637>
- Dekker J (2024) ZooMS identification of a whale bone tool from Bronze Age Heiloo, the Netherlands [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.10970629>
- Dekker J, Sinet-Mathiot V, Spithoven M, Smit B, Wilcke A, Welker F, Verpoorte A, Soressi M (2021) Human and cervid osseous materials used for barbed point manufacture in Mesolithic Doggerland. *Journal of archaeological science, reports*, 35, 102678. <https://doi.org/10.1016/j.jasrep.2020.102678>
- Desmond A, Barton N, Bouzouggar A, Douka K, Fernandez P, Humphrey L, Morales J, Turner E, Buckley M (2018) ZooMS identification of bone tools from the North African Later Stone Age. *Journal of archaeological science*, 98, 149–157. <https://doi.org/10.1016/j.jas.2018.08.012>
- van Doorn NL, Hollund H, Collins MJ (2011) A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. *Archaeological and anthropological sciences*, 3, 281. <https://doi.org/10.1007/s12520-011-0067-y>

- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, 356, 92–95. <https://doi.org/10.1126/science.aal3327>
- Edmonds K (2018) A bone to pick. A functional analysis of osseous materials from Bronze age West Frisia. MSc Thesis. Leiden University.
- Essel E, Zavala EI, Schulz-Kornas E, Kozlikin MB, Fewlass H, Vernot B, Shunkov MV, Derevianko AP, Douka K, Barnes I, Soulier M-C, Schmidt A, Szymanski M, Tsanova T, Sirakov N, Endarova E, McPherron SP, Hublin J-J, Kelso J, Pääbo S, Hajdinjak M, Soressi M, Meyer M (2023) Ancient human DNA recovered from a Palaeolithic pendant. *Nature*. <https://doi.org/10.1038/s41586-023-06035-2>
- Evans Z, Paskulin L, Rahemtulla F, Speller CF (2023) A comparison of minimally-invasive sampling techniques for ZooMS analysis of bone artifacts. *Journal of Archaeological Science: Reports*, 47, 103738. <https://doi.org/10.1016/j.jasrep.2022.103738>
- Évora MA (2015) Use-wear methodology on the analysis of osseous industries. In: *Use-Wear and Residue Analysis in Archaeology Manuals in Archaeological Method, Theory and Technique*. (eds Marreiros J, Gibaja Bao J, Ferreira Bicho N), pp. 159–170. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-08257-8_8
- Fahrner M, Kook L, Fröhlich K, Biniossek M, Schilling O (2021) A systematic evaluation of semispecific peptide search parameter enables identification of previously Undescribed N-terminal peptides and conserved proteolytic processing in cancer cell lines. *Proteomes*, 9. <https://doi.org/10.3390/proteomes9020026>
- Fontijn D (2002) *Sacrificial Landscapes: Cultural biographies of persons, objects and “natural” places in the Bronze Age of the southern Netherlands, c. 2300-600 BC*. Sidestone Press, Leiden, Netherlands.
- Fontijn D (2020) *Economies of Destruction How the systematic destruction of valuables created value in Bronze Age Europe, c. 2300-500 BC*. Routledge. <https://doi.org/10.4324/9781315109879>
- Foote AD, Kaschner K, Schultze SE, Garilao C, Ho SYW, Post K, Higham TFG, Stokowska C, van der Es H, Embling CB, Gregersen K, Johansson F, Willerslev E, Gilbert MTP (2013) Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nature communications*, 4, 1677. <https://doi.org/10.1038/ncomms2714>
- Gibb S (2024) MALDIquantForeign: Import/Export Routines for “MALDIquant.” <https://doi.org/10.32614/CRAN.package.MALDIquantForeign>
- Gibb S, Strimmer K (2012) MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, 28, 2270–2271. <https://doi.org/10.1093/bioinformatics/bts447>
- Goude G (2024) Prehistoric whaling and tool industry evidenced by advanced proteomic methods. *Peer Community in Archaeology*, 100501. <https://doi.org/10.24072/pci.archaeo.100501>
- Hansen J, Dekker J, Troché G, Fagernäs Z, Olsen JV, Saña Seguí M, Welker F (2024) A comparative study of commercially available, minimally invasive, sampling methods on Early Neolithic humeri analysed via palaeoproteomics. *Journal of archaeological science*, 167, 106002. <https://doi.org/10.1016/j.jas.2024.106002>
- Hansen J, Sierra A, Mata S, Gassiot Ballbé E, Rey Lanaspa J, Welker F, Saña Seguí M, Clemente Conte I (2024) Combining traceological analysis and ZooMS on Early Neolithic bone artefacts from the Cave of Coro Trasito, NE Iberian Peninsula: Cervidae used equally to Caprinae. *PloS ONE*. <https://doi.org/10.1371/journal.pone.0306448>
- van Hees E, van Amerongen Y (2023) Hout. In: *Opgraving Heiloo-Zuiderloo; Deelplan 3 & Middenduin Bewoningsresten uit de bronstijd en sporen uit de Romeinse tijd tot en met de Nieuwe tijd op de oostelijke flank van de strandwal in Heiloo, gemeente Heiloo Archol Rapport*. (ed Brattinga JJ), pp. 93–97.
- van der Heiden MJ (2018) *Bronstijdbewoning op de strandwal. Definitief Archeologisch onderzoek in het plangebied Zuiderloo, UWP 1, gemeente Heiloo (Noord-Holland)*. Diachron publicatie.

- Hennius A, Gustavsson R, Ljungkvist J, Spindler L (2018) Whalebone Gaming Pieces: Aspects of Marine Mammal Exploitation in Vendel and Viking Age Scandinavia. *European Journal of Archaeology*, 21, 612–631. <https://doi.org/10.1017/eea.2018.15>
- Hennius A, Ljungkvist J, Ashby SP, Hagan R, Presslee S, Christensen T, Gustavsson R, Peets J, Maldre L (2023) Late Iron Age Whaling in Scandinavia. *Journal of Maritime Archaeology*, 18, 1–22. <https://doi.org/10.1007/s11457-022-09349-w>
- Hofreiter M, Pacher M (1997) Using Ancient DNA to Elucidate Raw Material Origin of Bone Points from Potočka zijalka (Slovenia): Preliminary Results. *Mitt. Komm. Quatärforsch. Österr. Akad. Wiss.*, 13, 201–210.
- van den Hurk Y, Sikström F, Amkreutz L, Bleasdale M, Borvon A, Ephrem B, Fernández-Rodríguez C, Gibbs HMB, Jonsson L, Lehouck A, Cedeira JM, Meng S, Monge R, Moreno M, Nabais M, Nores C, Pis-Millán JA, Riddler I, Schmölcke U, Segschneider M, Speller C, Vretemark M, Wickler S, Collins M, Nadeau M-J, Barrett JH (2023) The prelude to industrial whaling: identifying the targets of ancient European whaling using zooarchaeology and collagen mass-peptide fingerprinting. *Royal Society open science*, 10, 230741. <https://doi.org/10.1098/rsos.230741>
- van den Hurk Y, Spindler L, McGrath K (2022) Medieval whalers in the Netherlands and Flanders: zooarchaeological analysis of medieval cetacean remains. *Environmental Archaeology*, 27. <https://doi.org/10.1080/14614103.2020.1829296>
- van den Hurk Y, Spindler L, McGrath K, Speller C (2020) The medieval mammoth: Biomolecular identification of mammoth remains from a Dutch medieval context. *Lutra*, 63, 75–79.
- Karunaratne P, Pocquet N, Labbé P, Milesi P (2022) BioRssay: an R package for analyses of bioassays and probit graphs. *Parasites & Vectors*, 15. <https://doi.org/10.1186/s13071-021-05146-x>
- Katoh K, Misawa K, Kuma K-I, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30, 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kenney RD (2009) Right Whales: *Eubalaena glacialis*, *E. japonica*, and *E. australis*. In: *Encyclopedia of Marine Mammals* (Second Edition) (eds Perrin WF, Würsig B, Thewissen JGM), pp. 962–972. Academic Press, London. <https://doi.org/10.1016/B978-0-12-373553-9.00220-0>
- Kirby DP, Buckley M, Promise E, Trauger SA, Holdcraft TR (2013) Identification of collagen-based materials in cultural heritage. *The Analyst*, 138, 4849–4858. <https://doi.org/10.1039/c3an00925d>
- Kleijne JP (2015) Kennemerland in de Bronstijd. *Provinciale Archeologische Rapporten*, 1.
- de Koning J, Tuinman NC (2019) Heiloo-Zuiderloo Riolsleuf. Sporen uit de Romeinse tijd, midden en late bronstijd. *Hollandia reeks*.
- Koning J, Tuinman NC (2021) De archeologische begeleiding van de aanleg van twee paddenpoelen bij Middenduin in Heiloo. Vroege middeleeuwen, Romeinse tijd en de bronstijd in een klein deel van Heiloo-Zuiderloo. *Hollandia reeks*.
- Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, 356. <https://doi.org/10.1186/s12859-014-0356-4>
- Lange S (2021) Hout. In: *Van kabeljauw tot karrenwiel. Bewoningsresten uit de bronstijd tot Romeinse tijd op de strandwal in plangebied Zuiderloo te Heiloo (proefsleuven en opgraving)* Archol rapport. (eds Moesker TP, van der Leije J, Tol AJ), p. 115.

- LeMoine GM (1994) Use wear on bone and antler tools from the Mackenzie Delta, Northwest Territories. *American antiquity*, 59, 316–334. <https://doi.org/10.2307/281935>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Mackie M, R  ther P, Samodova D, Di Gianvincenzo F, Granzotto C, Lyon D, Pegg   DA, Howard H, Harrison L, Jensen LJ, Olsen JV, Cappellini E (2018) Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angewandte Chemie*, 57, 7369–7374. <https://doi.org/10.1002/anie.201713020>
- Martisius NL, Welker F, Dogand  i   T, Grote MN, Rendu W, Sinet-Mathiot V, Wilcke A, McPherron SJP, Soressi M, Steele TE (2020) Non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals. *Scientific reports*, 10, 7746. <https://doi.org/10.1038/s41598-020-64358-w>
- McGowen MR, Tsagkogeorga G,   lvarez-Carretero S, Dos Reis M, Struebig M, Deaville R, Jepson PD, Jarman S, Polanowski A, Morin PA, Rossiter SJ (2020) Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. *Systematic biology*, 69, 479–501. <https://doi.org/10.1093/sysbio/syz068>
- McGrath K, Rowsell K, Gates St-Pierre C, Tedder A, Foody G, Roberts C, Speller C, Collins M (2019) Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points. *Scientific reports*, 9, 11027. <https://doi.org/10.1038/s41598-019-47299-x>
- Moesker TP, van der Leije J, Tol AJ (2021) Van Kabeljauw tot Karrewiel. Bewoningsresten uit de bronstijd tot Romeinse tijd op de strandwal in het plangebied Zuiderloo te Heiloo (proefsleuven en opgraving). Archol rapport.
- Morin PA, Alexander A, Blaxter M, Caballero S, Fedrigo O, Fontaine MC, Foote AD, Kuraku S, Maloney B, McCarthy ML, McGowen MR, Mountcastle J, Nery MF, Olsen MT, Rosel PE, Jarvis ED (2020) Building genomic infrastructure: Sequencing platinum-standard reference-quality genomes of all cetacean species. *Marine Mammal Science*, 36, 1356–1366. <https://doi.org/10.1111/mms.12721>
- Multari DH, Sullivan GJ, Hartley M, Power RK, Haynes PA (2023) Species identification of early colonial bone artefacts excavated from Pymont, Australia, by mass spectrometric identification of collagen peptides. *Journal of Archaeological Science: Reports*, 47, 103740. <https://doi.org/10.1016/j.jasrep.2022.103740>
- Mylopotamitaki D, Harking FS, Taurozzi AJ, Fagern  s Z, Godinho RM, Smith GM, Weiss M, Sch  ler T, McPherron SP, Meller H, Cascalheira J, Bicho N, Olsen JV, Hublin J-J, Welker F (2023) Comparing extraction method efficiency for high-throughput palaeoproteomic bone species identification. *Scientific reports*, 13, 18345. <https://doi.org/10.1038/s41598-023-44885-y>
- Nieweg D (2023) Botmateriaal. In: Opgraving Heiloo-Zuiderloo; Deelplan 3 & Middenduyn Bewoningsresten uit de bronstijd en sporen uit de Romeinse tijd tot en met de Nieuwe tijd op de oostelijke flank van de strandwal in Heiloo, gemeente Heiloo Archol rapport. (ed Brattinga JJ), pp. 75–86.
- Or  wska J, Cyrek K, Kaczmarczyk GP, Migal W, Osipowicz G (2023) Rediscovery of the Palaeolithic antler hammer from Bi  nik Cave, Poland: New insights into its chronology, raw material, technology of production and function. *Quaternary international: the journal of the International Union for Quaternary Research*, 665–666, 48–64. <https://doi.org/10.1016/j.quaint.2022.08.011>
- Pal Chowdhury M, Buckley M (2022) Trends in deamidation across archaeological bones, ceramics and dental calculus. *Methods*, 200, 67–79. <https://doi.org/10.1016/j.ymeth.2021.08.004>

- Palomo IR, Nair B, Chiang Y, Dekker J, Dartigues B, Mackie M, Evans M, Macleod R, Olsen JV, Collins MJ (2023) Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *bioRxiv*, 2023.12.15.571577. <https://doi.org/10.1101/2023.12.15.571577>
- Parker CE, Bos KI, Haak W, Krause J (2021) Optimized Bone Sampling Protocols for the Retrieval of Ancient DNA from Archaeological Remains. *Journal of visualized experiments: JoVE*. <https://doi.org/10.3791/63250>
- Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaíno JA (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic acids research*, 50, D543–D552. <https://doi.org/10.1093/nar/gkab1038>
- Presslee S (2020) ZooMS Markers: Published Data. Department of Archaeology, University of York.
- Procopio N, Chamberlain AT, Buckley M (2017) Intra- and Interskeletal Proteome Variations in Fresh and Buried Bones. *Journal of proteome research*, 16, 2016–2029. <https://doi.org/10.1021/acs.jproteome.6b01070>
- Ramsøe A, van Heekeren V, Ponce P, Fischer R, Barnes I, Speller C, Collins MJ (2020) DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *Journal of archaeological science*, 115, 105080. <https://doi.org/10.1016/j.jas.2020.105080>
- Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*, 2, 1896–1906. <https://doi.org/10.1038/nprot.2007.261>
- Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Ramsey CB, Butzin M, Cheng H, Lawrence Edwards R, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kromer B, Manning SW, Muscheler R, Palmer JG, Pearson C, van der Plicht J, Reimer RW, Richards DA, Marian Scott E, Southon JR, Turney CSM, Wacker L, Adolphi F, Büntgen U, Capano M, Fahrni SM, Fogtmann-Schulz A, Friedrich R, Köhler P, Kudsk S, Miyake F, Olsen J, Reinig F, Sakamoto M, Sookdeo A, Talamo S (2020) The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon*, 62, 725–757. <https://doi.org/10.1017/RDC.2020.41>
- Rey-Iglesia A, Martínez-Cedeira J, López A, Fernández R, Campos PF (2018) The genetic history of whaling in the Cantabrian Sea during the 13th–18th centuries: Were North Atlantic right whales (*Eubalaena glacialis*) the main target species? *Journal of Archaeological Science: Reports*, 18, 393–398. <https://doi.org/10.1016/j.jasrep.2018.01.034>
- Richards R (2009) Past and present distributions of southern right whales (*Eubalaena australis*). *New Zealand journal of zoology*, 36, 447–459. <https://doi.org/10.1080/03014223.2009.9651477>
- de Ridder T (1993) Stropers op de kust. Een prehistorische walvisstranding bij het zeegat van Bergen. *Westerheem*, 42, 57–61.
- Rüther PL, Husic IM, Bangsgaard P, Gregersen KM, Pantmann P, Carvalho M, Godinho RM, Friedl L, Cascalheira J, Taurozzi AJ, Jørkov MLS, Benedetti MM, Haws J, Bicho N, Welker F, Cappellini E, Olsen JV (2022) SPIN enables high throughput species identification of archaeological bone by proteomics. *Nature communications*, 13, 2458. <https://doi.org/10.1038/s41467-022-30097-x>
- Sawafuji R, Cappellini E, Nagaoka T, Fotakis AK, Jersie-Christensen RR, Olsen JV, Hirata K, Ueda S (2017) Proteomic profiling of archaeological human bone. *Royal Society open science*, 4, 161004. <https://doi.org/10.1098/rsos.161004>
- Sinet-Mathiot V, Martisius NL, Schulz-Kornas E, van Casteren A, Tsanova TR, Sirakov N, Spasov R, Welker F, Smith GM, Hublin J-J (2021) The effect of eraser sampling for proteomic analysis on Palaeolithic bone surface microtopography. *Scientific reports*, 11, 23611. <https://doi.org/10.1038/s41598-021-02823-w>

- Sinet-Mathiot V, Smith GM, Romandini M, Wilcke A, Peresani M, Hublin J-J, Welker F (2019) Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy). *Scientific reports*, 9, 12350. <https://doi.org/10.1038/s41598-019-48706-z>
- Speller C, van den Hurk Y, Charpentier A, Rodrigues A, Gardeisen A, Wilkens B, McGrath K, Rowsell K, Spindler L, Collins M, Hofreiter M (2016) Barcoding the largest animals on Earth: ongoing challenges and molecular solutions in the taxonomic identification of ancient cetaceans. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371. <https://doi.org/10.1098/rstb.2015.0332>
- Surovell TA, Litynski ML, Allaun SA, Buckley M, Schoborg TA, Govaerts JA, O'Brien MJ, Pelton SR, Sanders PH, Mackie ME, Kelly RL (2024) Use of hare bone for the manufacture of a Clovis bead. *Scientific reports*, 14, 2937. <https://doi.org/10.1038/s41598-024-53390-9>
- Tejero J-M, Cheronet O, Gelabert P, Zagorc B, Álvarez-Fernández E, Arias P, Averbouh A, Bar-Oz G, Barzilai O, Belfer-Cohen A, Bosch MD, Brück F, Cueto M, Dockner M, Fullola JM, Gárate D, Giannakoulis M, González C, Jakeli N, Mangado X, Meshveliani T, Neruda P, Nigst P, Ontañón R, Shemer M, Šimková PG, Tapia J, Sánchez de la Torre M, Schwab C, Weber G, Pinhasi R (2024) Cervidae antlers exploited to manufacture prehistoric tools and hunting implements as a reliable source of ancient DNA. *Heliyon*, 10, e31858. <https://doi.org/10.1016/j.heliyon.2024.e31858>
- Tuinman NC, de Koning J, Vermoolen S, Tol M (2022) Heiloo-Krommelaan/Zandbergen. *Archeologische Kroniek van Noord-Holland*, 2021, 79–80.
- Van Gijn A (2007) The use of bone and antler tools: two examples from the Late Mesolithic in the Dutch Coastal Zone. In: *Bones as Tools: Current Methods and Interpretations in Worked Bone Studies* (eds Gates St. Pierre C, Walker RB), pp. 81–92. BAR Publishing, Oxford, England.
- Verbaas A, Edmonds K (2018) Gebruikssporenanalyse van werktuigen uit been en gewei gevonden bij de opgravingen van de N23 West Frisiaweg. Lab.
- Welker F, Hajdinjak M, Talamo S, Jaouen K, Dannemann M, David F, Julien M, Meyer M, Kelso J, Barnes I, Brace S, Kamminga P, Fischer R, Kessler BM, Stewart JR, Pääbo S, Collins MJ, Hublin J-J (2016) Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 11162–11167. <https://doi.org/10.1073/pnas.1605834113>
- Welker F, Soressi M, Rendu W, Hublin J-J, Collins M (2015) Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *Journal of archaeological science*, 54, 279–286. <https://doi.org/10.1016/j.jas.2014.12.010>
- Wellman HP, Rick TC, Rodrigues AT, Yang DY (2017) Evaluating Ancient Whale Exploitation on the Northern Oregon Coast Through Ancient DNA and Zooarchaeological Analysis. *The Journal of Island and Coastal Archaeology*, 12, 255–275. <https://doi.org/10.1080/15564894.2016.1172382>
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. *Journal of open source software*, 4, 1686. <https://doi.org/10.21105/joss.01686>
- Wolf M, Zapf K, Gupta DK, Hiller M, Árnason Ú, Janke A (2023) The genome of the pygmy right whale illuminates the evolution of rorquals. *BMC biology*, 21, 79. <https://doi.org/10.1186/s12915-023-01579-1>
- Zeiler J, Rijkelijkhuizen M, van Dijk J, Esser K Bone and antler tools in The Netherlands. (forthcoming) Rijksdienst voor het Cultureel Erfgoed.

3 An unclean slate, discrepancies between food input and recovered protein signal from experimental foodcrusts.

Joannes Adrianus Antonius Dekker^{1,2}, Richard Hagan², Matthew James Collins^{1,3}, Jessica Hendy².

1: Section for Geobiology, Globe Institute, University of Copenhagen, Copenhagen, Denmark.

2: BioArCh, Department of Archaeology, University of York, York, UK.

3: McDonald Institute for Archaeological Research, University of Cambridge, UK.

This chapter was submitted to the Royal Society Open Science journal on 17-02-2025. Additionally, the chapter is available as a preprint on bioRxiv under the following identifier: <https://doi.org/10.1101/2025.02.03.636240>.

CRedit statement:

J.A.A.D: Conceptualisation, Methodology, Validation, Formal analysis, Investigation, Investigation, Writing - Original draft, Writing - Review & editing, Visualisation, Project administration. **M.C:**Resources, Writing - Review & editing, Supervision, Project administration, Funding acquisition. **J.H:**Resources, Writing - Review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements:

We thank the Bioscience Technology Facility and Chemistry Department at the University of York for mass spectrometry access and support, in particular we would like to thank Dr. Adam Dowle for his support in running the samples and feedback on the manuscript. The instrumentation is part of the York Centre of Excellence in Mass Spectrometry. The centre was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1).

We would like to thank the YEAR Centre for their permission to use the Centre's grounds for the burial experiment and in particular Dr. Gareth Perry for his assistance in the burial and excavation of the experimental vessels.

For the morphological taxonomic verification of the *Salmo salar* specimens, we would like to thank Dr. Katrien Dierickx, and for his assistance in figuring out Sankey diagrams we thank Jakob Hansen. Additionally, we would like to thank Dr. Claire Koenig for her advice on protein quantification.

Additionally, our gratitude goes to Patrick Gibbs and James Osborn from Heritage360, University of York, for access to their 3D scanning equipment and assistance in the acquisition and analysis of the 3D data.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956351.

Abstract

Organic residues are a rich source of biomolecular information on ancient diets. In particular, foodcrusts, charred residues on ceramics, are commonly analysed for their lipid content and to a lesser extent protein in order to identify foods, culinary practices and material culture use in past populations. However, the composition of foodcrusts and the factors behind their formation are not well understood. Here we analyse lipid and proteomic data from foodcrusts made using a series of mixtures of protein- (salmon flesh), lipid- (beef fat) and carbohydrate-rich (beetroot) foods to investigate the relationship between the biomolecular composition of the input and the recovered signal using conventional methods applied to archaeological material. Additionally, using 3D modelling we quantify the volume of foodcrust generated by different ingredient combinations. The results highlight biases in the data obtained in the analyses of organic residues both in terms of identified resources reflecting the cooked foodstuffs, e.g. an overrepresentation of fish proteins, as well as with regards to the abundance of foodcrust, for example mixtures of only salmon and beef fat resulted in relatively small amounts of foodcrust, and suggest caution in interpreting the composition of residues formed from complex mixtures of foodstuffs.

3.1 Background

Food residues on or in ceramic vessels are a popular target for studies on ancient diets and foodways. Their potential for revealing components of the diets of past people is demonstrated by studies on the adoption of pottery by hunter-gatherers (Oras et al. 2017) and continuity in diet under political regime change (Lundy et al. 2023). Organic residues on archaeological vessels can manifest in different ways, including calcified deposits (Hendy et al. 2018b), charred foodcrusts (hereafter foodcrusts) (Shevchenko et al. 2018) and organic residues entrapped within the ceramic matrix (Pal Chowdhury, Campbell and Buckley 2021), and more rarely, organic remains preserved on the surface of vessels through arid, frozen or waterlogged conditions (Cronin et al. 2007; Greco et al. 2018; Yang et al. 2014). These three different types of residues are formed by different processes and are affected differently by diagenesis. Of the three types of residue we focus on foodcrusts in this study. Foodcrusts have so far been mostly analysed for their lipid content (including fats, oils and waxes). Lipid analysis is robust and has been successfully applied to materials from a wide range of preservation conditions spanning the globe and stretching back as far as at least 12 000 years ago (Craig et al. 2013; Admiraal et al. 2020, 2023; Lundy et al. 2023). For example, the study of lipids has been successful in studying the emergence and spread of the use of dairy products (Copley et al. 2003, 2005; Craig et al. 2005) as well as the continuity in the widespread use of marine resources through the Mesolithic into the Neolithic (Craig et al. 2011; Bondetti et al. 2020; Courel et al. 2020; Lucquin et al. 2023). However, there are still areas for improvement. The results of compound-specific isotope analysis can be confounded by the mixture of multiple resources and the results are likely biased towards lipid-rich foodstuffs compared to lipid-poor foods, such as most plants (Colonese et al. 2017).

The presence of plants in foodcrusts is best investigated using SEM (*Scanning Electron Microscopy*) (Courel et al. 2021; Kubiak-Martens et al. 2024), revealing, for example, the presence of acorns in ceramic vessels in the Neolithic Netherlands (Kubiak-Martens, Brinkkemper and Oudemans 2015). Animal remains, although difficult to taxonomically resolve, can also be observed using SEM (Courel et al. 2021).

In theory, the disadvantages of both lipid analysis and SEM could be circumvented by protein analysis. As proteins are encoded by DNA they inherently carry phylogenetic information, although the taxonomic specificity of a protein varies greatly between proteins. Beyond their use for taxonomic identification, some proteins can be used to identify the tissues that were cooked, as some proteins are abundant in or exclusive to specific tissues. For example, the frequently observed protein *beta-lactoglobulin* (BLG) is specific to dairy (Warinner et al. 2014; Charlton et al. 2019; Bleasdale et al. 2021; Evans et al. 2023), and is not synthesised by humans at all. Additionally, as proteomic taxonomic identification is based on the presence of diagnostic peptide sequences, protein analysis should be able to distinguish between the different components of a food mixture. However, the application of protein analysis to foodcrusts is in its infancy (Solazzo et al. 2008; Greco et al. 2018; Shevchenko et al. 2018; Tanasi et al. 2021). While some success has been observed (Shevchenko et al. 2018; Lyu et al. 2024), biases and challenges in this method have been noted [12]. For example, the cooking and burning process of foodcrust formation likely impacts the numbers of food proteins, which may or may not be taxonomically diagnostic. Furthermore, in the case of analysing mixtures of ingredients, it is unknown whether the foodcrust preserves an accurate reflection of the input food proteins. Protein analysis will of course favour protein-rich foods, but other foodstuffs less enriched in protein, such as plants, still contain characteristic proteins and have been identified in organic residues (Hendy et al. 2018b). Whether there are additional factors beyond the initial protein content determining the likelihood of a particular protein's survival is unclear. A recent study on protein incorporation and survival in the ceramic matrix and foodcrust tested the influence of several physico-chemical properties, but beyond a slight favour towards hydrophobic peptides, there were no clear universal trends in protein survival (Evans et al. 2024). Beyond these general limitations, not much is known about the challenges of applying protein analysis to organic residues nor about how the theoretical advantages translate into reality.

This study will focus on foodcrusts and aims to clarify the advantages and disadvantages of protein analysis compared to lipid analysis using experimental foodcrusts made of a mixture of protein-, lipid- and carbohydrate-rich foods. Specifically, we will investigate how mixtures of different foodstuffs might influence the volume of foodcrust formed and its protein and lipid content. In particular, we will examine the degree to which the 'extractome', the suite of recovered proteins, (Schroeter et al. 2016) reflect the original input food mixture, in order to highlight the potential biases involved in the identification of foodstuffs in archaeological material. Additionally, we will investigate how repeated cooking events and burial influence the extractome's composition and the volume of generated foodcrust, adding to our understanding of the diagenesis of proteins in charred organic residues.

In order to maximise the relevance of the experiment's results for archaeological samples the choice of foodstuffs is vital. As this experiment is part of a larger project focused on Mesolithic foodways in Denmark, we selected foodstuffs that would be similar to those utilised by late Danish hunter-gatherers. Secondly, as lipid analysis favours lipid-rich resources and protein analysis protein-rich foodstuffs, we decided to select one lipid-rich, one protein-rich and one carbohydrate-rich foodstuff to mix. Using foodstuffs with vastly

different nutritional compositions facilitates visualising the biases towards either carbohydrates, proteins or lipids. Therefore, care was taken to ensure that the relative proportions of the wet weights of the carbohydrate-rich, protein-rich and lipid-rich foodstuffs matched as closely as possible to the ratio of carbohydrates:proteins:lipids in the total mixture (e.g. a mixture consisting of 20% of the protein-rich food would also contain 20% proteins). Perfect symmetry between the ratios of wet weights and the molecular categories was not possible, especially not when limited to foodstuffs likely available in Mesolithic Denmark, but the closest match was obtained using salmon, beef fat and beetroot (SI 1). However, we acknowledge that this combination of foodstuffs underrepresents the carbohydrate component in particular.

Beetroot was chosen as the carbohydrate-rich food due to the prevalence of its wild ancestor *Beta vulgaris maritima* at Mesolithic sites such as Tybrind Vig and Møllegabet II (Kubiak-Martens 1999; Mason 2004), as well as the fact that it has a relatively well-annotated reference proteome. Protein identification is heavily dependent on the quality and coverage of reference proteomes. Low-quality reference proteomes only contain a fraction of the total proteome of a species and any proteins not included cannot be identified. To highlight the magnitude of this issue, the median number of proteins listed on UniProt (UniProt Consortium 2023) for 29 plant species found at Mesolithic northern European sites (Kubiak-Martens 1999, 2002; Price et al. 2001; Holst 2010; Robinson and Harild 2009) is 120 proteins (SI 2), in contrast to the 7679 available for *Beta vulgaris* or the 130 673 for *Triticum aestivum* (wheat), a high-quality reference proteome.

We selected beef fat and salmon because the exploitation of aurochs (*Bos primigenius*) and salmon (*Salmo salar*) is well attested at a number of Danish late Mesolithic sites (Price et al. 2001; Richter and Noe-Nygaard 2003; Astrup et al. 2021; Enghoff 2011; Price et al. 2018). Additionally, salmon has a well-annotated reference proteome covering 90% of its total proteome. Available reference data for aurochs is much more limited, but due to the short evolutionary distance between them, domestic cattle (*Bos taurus*), which has a 97% complete reference proteome, was used.

3.2 Materials

The ceramic vessels used in the cooking experiments were purchased from Potted History, a company specialising in historical ceramic replicas. The specifications of the ceramics are in line with the pottery used in the experiments by Bondetti et al. (Bondetti et al. 2021): they were hand thrown with the 'standard red' clay (Al_2O_3 , 22.78; Fe_2O_3 , 7.37; CaO , 0.57; MgO , 0.86; K_2O , 1.6; and Na_2O , 0.1), had a diameter of ~10 cm, a wall thickness of ~0.5 cm and were fired at 700 °C. The foodstuffs used in the experiments were all obtained from local commercial vendors; the beetroot was purchased from Morrisons, the beef fat from M&K Quality Butchers, York, and the salmon was purchased online from LondonGrocery.net (<https://londongrocery.net/>). As what is often sold as 'salmon' can actually belong to a variety of species and even genera, the purchased specimens were examined for morphological characteristics (Page and Burr 2011) to ensure that they were in fact *Salmo salar*.

3.3 Methods

3.3.1 Cooking experimental protocol

In order to test how different combinations of carbohydrate-, protein- and lipid-rich foodstuffs influence foodcrust formation it was decided to mix the three selected foodstuffs, beetroot, salmon and beef fat, in all possible combinations using 20% increments, which results in 21 sample conditions (Fig. 1). To each of these unique combinations one ceramic vessel was assigned. Additionally, we included one ceramic vessel as a negative control, which would only be filled with water during cooking.

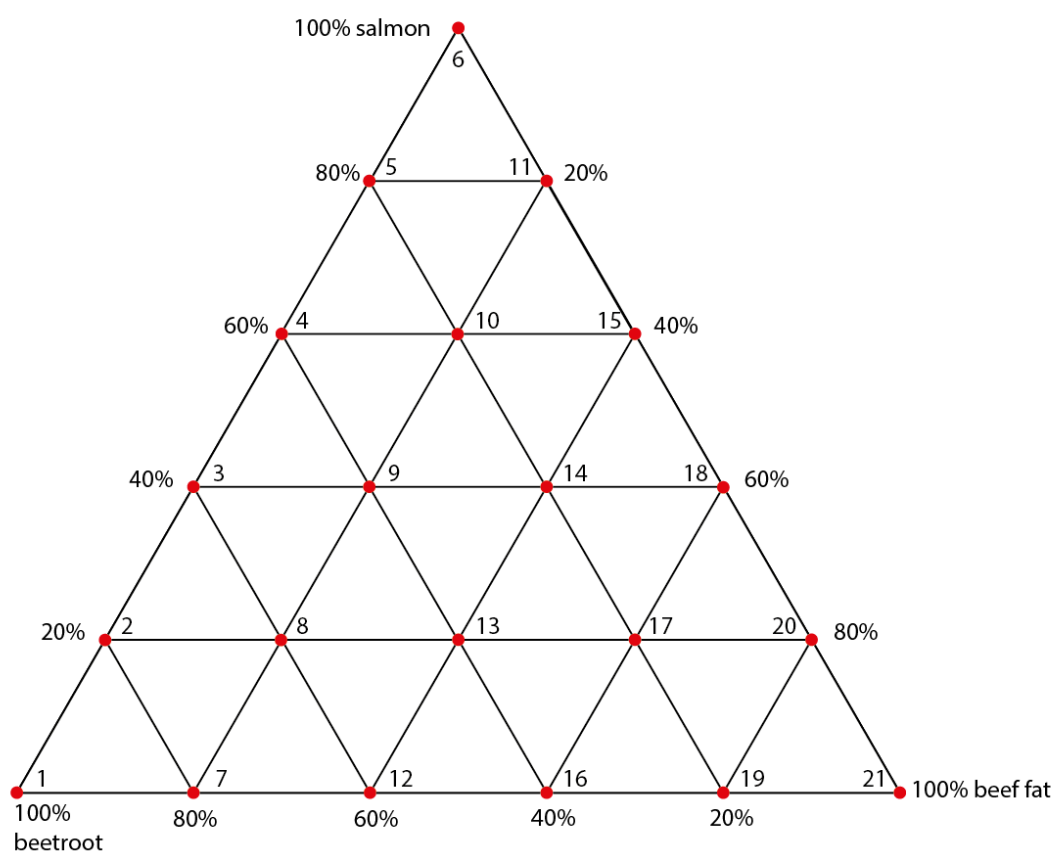


Figure 1, Overview of experimental conditions. Each red node indicates an experimental vessel for a total of 21 vessels.

Before the first cooking event, all pots were soaked in tap water at room temperature for 15 minutes in order to prevent cracking when heated. The vessels were then filled with the ingredients according to their experimental conditions to a total mass of 100 g. Additionally, 100 mL of distilled water was added to each vessel. The filled vessels were wrapped in aluminium foil, leaving the top open, and placed for an hour in an oven (Binder) preheated to 270°C. The aluminium foil was intended to hold the pots and their contents in place in the event of the ceramic fracturing, as pilot tests had shown this to be a possibility. The cooking temperature of 270°C was selected based on previous research that found that APAAs (ω -(O-alkylphenyl) alkanolic acids) often found in archaeological foodcrusts, formed

after heating for one hour at 270°C (Bondetti et al. 2021). After cooking for one hour, the vessels were removed from the oven and allowed to cool down at room temperature. The cooled foodstuffs were then removed from the vessels as we were interested in multiple cooking events. Care was taken to only remove the foodstuffs and to leave the foodcrusts intact. Charred remains that adhered to the ceramic wall, but were macroscopically still recognisable as food, were removed as well. Fig. 2 provides some examples of foodcrusts that were formed after the final cooking event.

1, 100% beetroot



6, 100% salmon



21, 100% beef fat



2, 80% beetroot, 20% salmon



13, 40% beetroot, 20% salmon, 40% beef fat



15, 60% salmon, 40% beef fat



Figure 2, Foodcrusts formed after the ninth cooking event. The number denotes the experimental of each vessel and corresponds to Fig. 1.

After removing the food the vessels were upturned for roughly an hour to allow any liquid to drip out. Any fat that consolidated on a vessel's rim during this period was not considered foodcrust and removed. For subsequent cooking episodes, the vessels were no longer soaked in water. Instead the vessels were immediately placed in the oven after filling them with their ingredients and 100 mL water. In total nine one-hour cooking events were performed for each pot and foodcrust samples were taken after the first, fifth and ninth cooking episodes. Foodcrust could form on the rim, body and bottom of the vessel

depending on the food input, but the rim of the vessel was the only location that always contained sufficient foodcrust for sampling, and hence foodcrust was consistently sampled from the rim. Where possible samples were still taken from the bottom and body of the vessel. No replicate samples were taken from any location or cooking event.

After the ninth cooking episode, the vessels were buried with their foodcrusts in the York Experimental Archaeological Research (YEAR) Centre for 52 days at a depth of 50 cm. The vessels were wrapped loosely in Galvanised wire netting (13 mm mesh size) to prevent bioturbation from any larger creatures. Insects and microbes are expected to still have access to the foodcrusts. After the burial period had passed, the vessels were excavated and a foodcrust sample was taken from each vessel. All samples were stored at -20°C until protein extraction.

3.3.2 Protein extraction

Foodcrust protein analysis was performed according to a variant of the SP3 protocol (Cleland 2018) developed by Dr. Virginia Harvey and Dr. Jessica Hendy (Harvey and Hendy 2025). The extraction was performed in batches and an extraction blank was included in every batch for a total of 11 blanks. In short, 150 µL 6 M guanidine hydrochloride (GuHCl) was added to 20 mg foodcrust from each of the sampled experimental conditions, locations and time points, homogenised, centrifuged and incubated at 65°C for one hour. Subsequently, the samples were centrifuged and the foodcrust pellets were discarded. 15 µL of a 100 mM TCEP (Tris(2-carboxyethyl)phosphine) and CAA (2-Chloroacetamide) solution was added to the supernatant, homogenised, centrifuged and incubated at 99°C for 10 minutes. After the incubation, the samples were allowed to cool in room temperature conditions, after which 500 µg of magnetic hydrophilic and hydrophobic beads (Cytiva) and 175 µL of 100% (v/v) ethanol were added to each sample. The samples were then incubated again at 24°C, 1000 RPM for 5 minutes. Protein filtration using the beads was performed by placing the samples on a magnetic rack and allowing the beads to migrate for 2 minutes. The liquid in the samples was then removed until only the beads remained. Samples were taken off the rack and were rinsed with 80% ethanol. This process was repeated three times, the first time 500 µL ethanol was used for the rinsing, the second time 300 µL and the third rinse was done with 200 µL. After the final rinse, the beads were resuspended in 100 µL 50 mM ammonium bicarbonate (AmBic). This solution was incubated at 37°C with 750 RPM for 3 minutes, after which the samples were placed back on the magnetic rack. In-solution digestion was performed by adding 1 µL of a 0.2 µg/µL trypsin solution to the samples and incubating them at 37°C, 750 RPM for 18 hours.

After digestion, the samples were centrifuged and placed back on the magnetic rack. Beads and suspended proteins were separated by transferring the solution to another set of Eppendorf tubes. 10 µL of 5% (v/v) trifluoroacetic acid (TFA) was added to the samples to halt digestion. As a last step, the peptides were desalted with C18 ZipTips. The ZipTips were primed twice with a 100 µL 50% acetonitrile (ACN) and 0.1% TFA solution and washed twice with 0.1% TFA. Then the peptides were aspirated from the sample and the ZipTips were washed again twice with 0.1% TFA before eluting the peptides in 50% ACN and 0.1% TFA.

3.3.3 Protein Mass Spectrometry analysis

The eluted peptides were analysed using an mClass nanoflow UPLC (Ultra Performance Liquid Chromatography) system (Waters) coupled to an Orbitrap Fusion Tribrid

mass spectrometer (Thermo Scientific) following the same protocol as published elsewhere (Dekker et al. 2023). In short, the UPLC was equipped with a nanoEase M/Z Symmetry 100 Å C18, 5 µm trap column (180 µm x 20 mm, Waters) and a PepMap, 2 µm, 100 Å, C18 EasyNano nanocapillary column (75 µm x 500 mm, Thermo Scientific). Chromatographic separation of the peptides was performed by first washing the peptides from the trap column onto the capillary column with 0.05% trifluoroacetic acid aqueous solvent at a flow rate of 15 µL/min. After five minutes the flow was switched to the capillary column, which was first washed for seven minutes with solvent A (0.1% FA (formic acid)) and 3-10% solvent B (0.1% FA in acetonitrile). The concentration of solution B in the mixture was steadily increased over time. After these initial seven minutes, the concentration of solvent B was increased to 35% over 30 minutes and lastly to 99% over five minutes. Finally, the column was washed with 99% solvent B for four minutes.

As the peptides were eluted off the column they were introduced into the mass spectrometer, which was equipped with an EasyNano ionisation source (Thermo Scientific). Xcalibur (version 4.0, Thermo Scientific) was used to acquire ESI-MS and MS² spectra both in positive mode. The following instrument source settings were used: ion spray voltage, 1,900 V; sweep gas, 0 Arb; ion transfer tube temperature; 275°C. MS¹ spectra were acquired in the Orbitrap with: resolving power: 120,000 *Full Width at Half Height* (FWHH), scan range: m/z 375-1,500; AGC target, 4e⁵; max fill time, 100 ms. Acquisition was data dependent (DDA) and performed in topN mode selecting the 12 most intense precursors with charge states >1. Easy-IC was used for internal calibration. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at 5e³. Precursors were fragmented using HCD. MS² spectra were acquired in the Orbitrap with: resolving power: 30,000 FWHH, max fill time, 100 ms., HCD; activation energy: 32 *Normalised Collision Energy* (NCE).

3.3.4 Protein data analysis

The resulting raw files were analysed with MetaMorpheus (Solntsev et al. 2018) using the following settings: files were calibrated with a product mass tolerance of 25 ppm and a precursor tolerance of 15 ppm. A G-PTM-D (Global-Post Translational Modification Discovery) task was performed in order to identify the most common post-translational modifications in the samples and for the search task a product mass tolerance of 0.5 Da and precursor mass tolerance of 3 ppm was maintained. The most commonly observed PTMs were oxidation of methionine, deamidation of asparagine, hydroxylation of proline and deamidation of glutamine. The maximum number of missed cleavages was 2 and the minimum peptide length was set to 7. Only fully tryptic peptides were accepted.

Using these settings the samples were searched against a custom database consisting of the cRAP (common Repository of Adventitious Proteins) contaminant database (The Global Proteome Machine Organization 2012) and the reference proteomes of *Beta vulgaris*, *Chenopodium quinoa*, *Spinacia oleracea*, *Salmo salar* and *Bos taurus*, which were downloaded from the Uniprot protein database (UniProt Consortium 2023). In addition to the proteomes of the species used in the cooking experiments, it was also decided to include quinoa (*Chenopodium quinoa*) and spinach (*Spinacia oleracea*), because of the limited coverage of the *Beta vulgaris* reference proteome. We feared that our analysis would bias against *Beta vulgaris* and therefore decided to include the two species with the largest reference proteomes from the Amaranthaceae family, which includes *Beta vulgaris*. The cRAP contaminant database consists of commonly found contaminants, such as human skin

proteins and protein standards. The cRAP database does not contain any soil bacteria. The exact task files and protein database used in this analysis are included in SI 3.

The MetaMorpheus protein identifications were then filtered following the recommendations for palaeoproteomic identifications (Hendy et al. 2018a), rejecting all proteins which were present in the blanks or which were represented by less than two unique peptides.

Statistical analysis was performed in R v4.1.1 (R Core Team 2021) using the Rstudio environment v2024.4.0.735 (Posit team 2024).

3.3.5 Foodcrust quantification

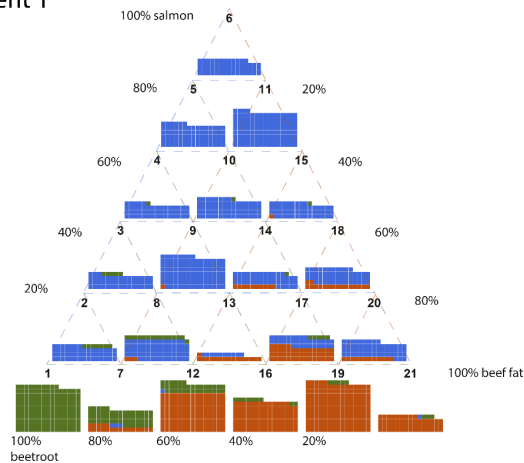
In order to investigate the influence of ingredient composition on the amount of foodcrust generated after cooking we quantified the volume of foodcrust using 3D scanning. A 3D scan of each ceramic vessel was acquired before the first cooking event and after the ninth using an Artec Space Spider (Artec 3D). 3D model construction and further analysis of the 3D models were all performed in Artec Studio 16 Professional v16.0.8.2. Volume measurements were obtained by combining the scans for each vessel and time point using the 'Sharp fusion' function. Then the volume of the entire 3D model was calculated using the in-built volume calculation function. The foodcrust volume was calculated by subtracting the before-cooking 3D model volume from the after-cooking 3D model volume.

Vessels 1, 7, 11 and 21 were damaged during the cooking process and some fragments had detached from the ceramic wall. To prevent the loss of ceramic fragments from influencing the foodcrust quantification the ceramic wall from which the fragment had detached was deleted from both the before and after model. The resulting hole in the 3D models was closed using the 'Fix holes' function. Irregularities in the 3D model resulting from this intervention were resolved using the 'Small-object filter' and 'Hole filling' 'Postprocessing' functions. Foodcrust volume calculations could then be obtained in the same way as for the undamaged experiments.

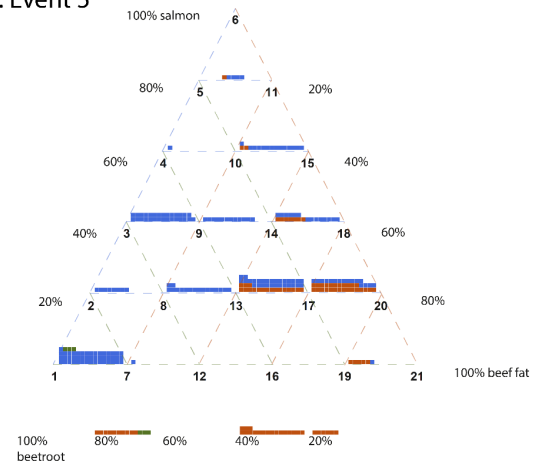
3.4 Results

A total list of all proteins recovered from the samples can be found in SI 4. There appeared to be substantial differences in the number of proteins that could be identified between the different experimental conditions and the number of cooking events for a particular ingredient combination (Fig. 3). Additionally, most samples show large differences in the contribution of the different ingredients to the total amount of recovered proteins. Some samples, 6.5, 7.1, 12.1, 12.9, 15.1 and 21.1, also yielded proteins from a species not included in their mixture (Table 1).

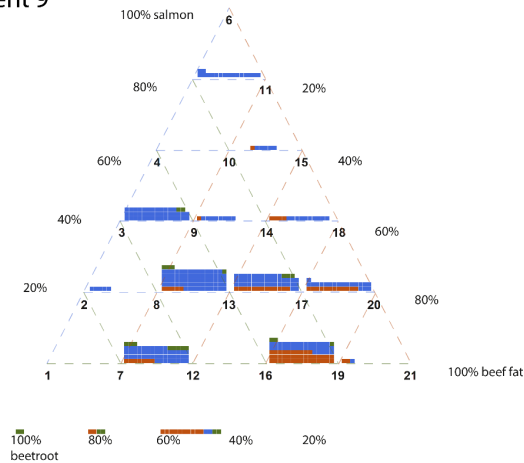
A: Event 1



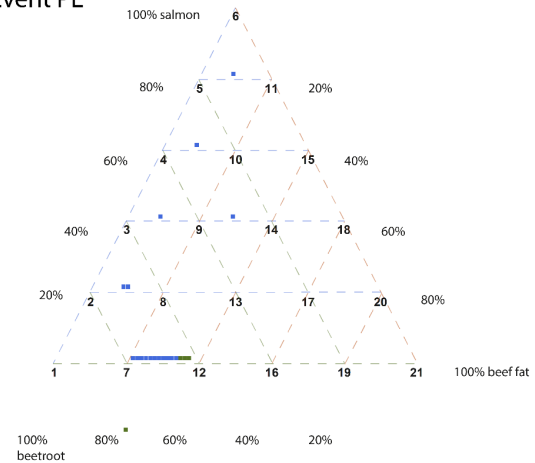
B: Event 5



C: Event 9



D: Event PE



Taxon

Beetroot

Salmon

Beef fat

Figure 3, Number of identified in each experimental condition after cooking event 1 (A), cooking event 5 (B), cooking event 9 (C) and post-excavation (D). Each square represents 1 identified protein and is coloured to reflect its taxonomic identity. *Beta vulgaris*, *Chenopodium quinoa* and *Spinacia oleracea* are grouped together as 'plant proteins'.

Table 1, Identified proteins belonging to taxa not included in the samples' mixture.

Experiment number	Cooking event	Protein Accession	Protein Full Name	Sequence Coverage	Number of Unique Peptides	Organism
6	5	A0A3Q1LL35	Uncharacterized protein	25%	2	Bos taurus
		A0A1S2X522	phosphopyruvate hydratase	63%	4	Salmo salar
7	1	B5DG55	Alpha-1,4 glucan phosphorylase	48%	5	Salmo salar
		Q3ZLR1	Superoxide dismutase [Cu-Zn]	23%	2	Salmo salar
	1	C0H9B8	Trifunctional enzyme subunit beta, mitochondrial	9%	2	Salmo salar
12	9	A0A1S3Q7E3	collagen alpha-1(I) chain-like	10%	2	Salmo salar
		A0A1S3SM10	collagen alpha-1(I) chain-like	8%	2	Salmo salar
15	1	Q9LEE0	phosphopyruvate hydratase	15%	2	Spinacia oleracea
		A0A0J7YLM1	Elongation factor Tu (Fragment)	13%	4	Beta vulgaris subsp. vulgaris
21	1	A0A0J8B1E8	50S ribosomal protein L5, chloroplastic	14%	2	Beta vulgaris subsp. vulgaris
		A0A0K9Q7T6	S1 motif domain-containing protein	4%	2	Spinacia oleracea

Fig. 3 also shows that for most experimental conditions the number of protein identifications decreases as the number of cooking events increases, which fits with the expected increase in protein damage due to extended heating. To further explore this pattern we plotted the distribution of protein identifications per sample point (Fig. 4). For each sample we also calculated the protein input in grams based on the ingredients' protein content (Public Health England 2021). As expected the number of proteins identified decreases after the first sampling point and is lowest after the burial, but some of the samples obtained after the ninth cooking event contain more proteins than any sample taken after the fifth event. To test if these differences are statistically significant, a Kruskal Wallis test (Kruskal and Wallis 1952) was performed ($p = 4.183 \times 10^{-8}$), which indeed suggests that there are significant differences between the number of proteins recovered after the different cooking events. Post hoc analysis using Dunn's test (Dunn 1961; Dinno 2017) shows that there are significant differences between cooking events 1 and 5 ($p = 5.46 \times 10^{-5}$), 1 and 9 ($p = 1.08 \times 10^{-3}$), 1 and PE (5.78×10^{-7}) and lastly between events 9 and PE ($p = 0.0497$). The seeming difference in protein recovery between events 5 and 9 is therefore not statistically significant.

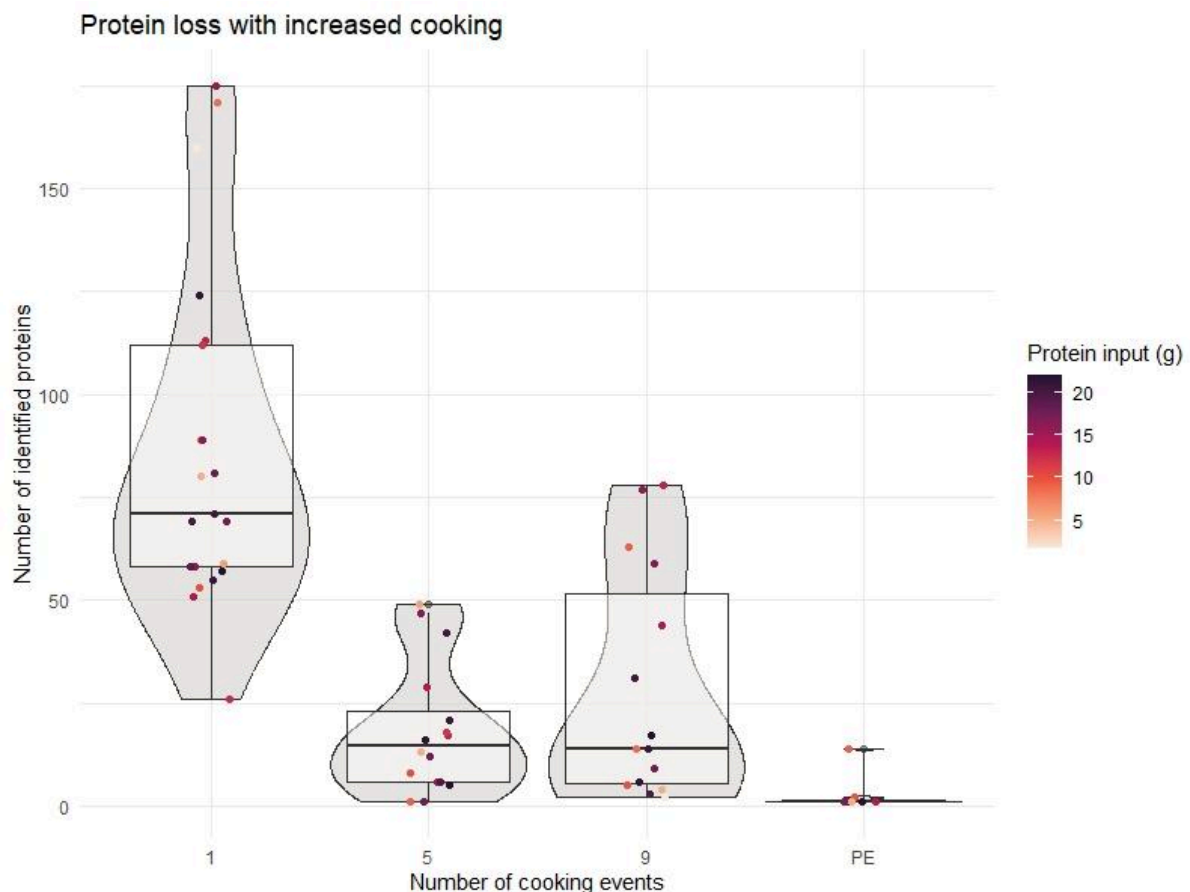


Figure 4, Distribution of the number of proteins identified per sampling point. In general, the number of identified proteins decreases with increasing cooking events, with two vessels acting as outliers.

To further investigate the relationship between the observed proteome abundance and the cooked ingredients we processed the identified peptides in the following way: first, we filtered the identified peptides to only include those that were observed in the single-species samples (i.e. experiment numbers 1, 6 and 21). As the number of identified

peptides, let alone shared-peptides, decreases rapidly with increased cooking, we only included samples from the first cooking event. For each protein in each sample, we then selected the three peptides with the highest intensities (Fig. 5) and summed their intensities per sample, giving a summed intensity as a proxy of their protein abundance. Using these summed intensities as the response variable we created a linear regression model with the nutritional composition as predictors. In the first model we split the fat content of each sample by saturated, mono-unsaturated and poly-unsaturated fatty acids, while in the second model, we grouped these three fatty acid subcategories together. Both models showed a significant relationship ($R^2 = 0.7765$, $F(2,18) = 35.75$), the total protein ($p = 1.16e^{-6}$) and carbohydrate ($p = 4.84e^{-6}$) content, in particular, showed a significant positive correlation with the observed sum of intensities.

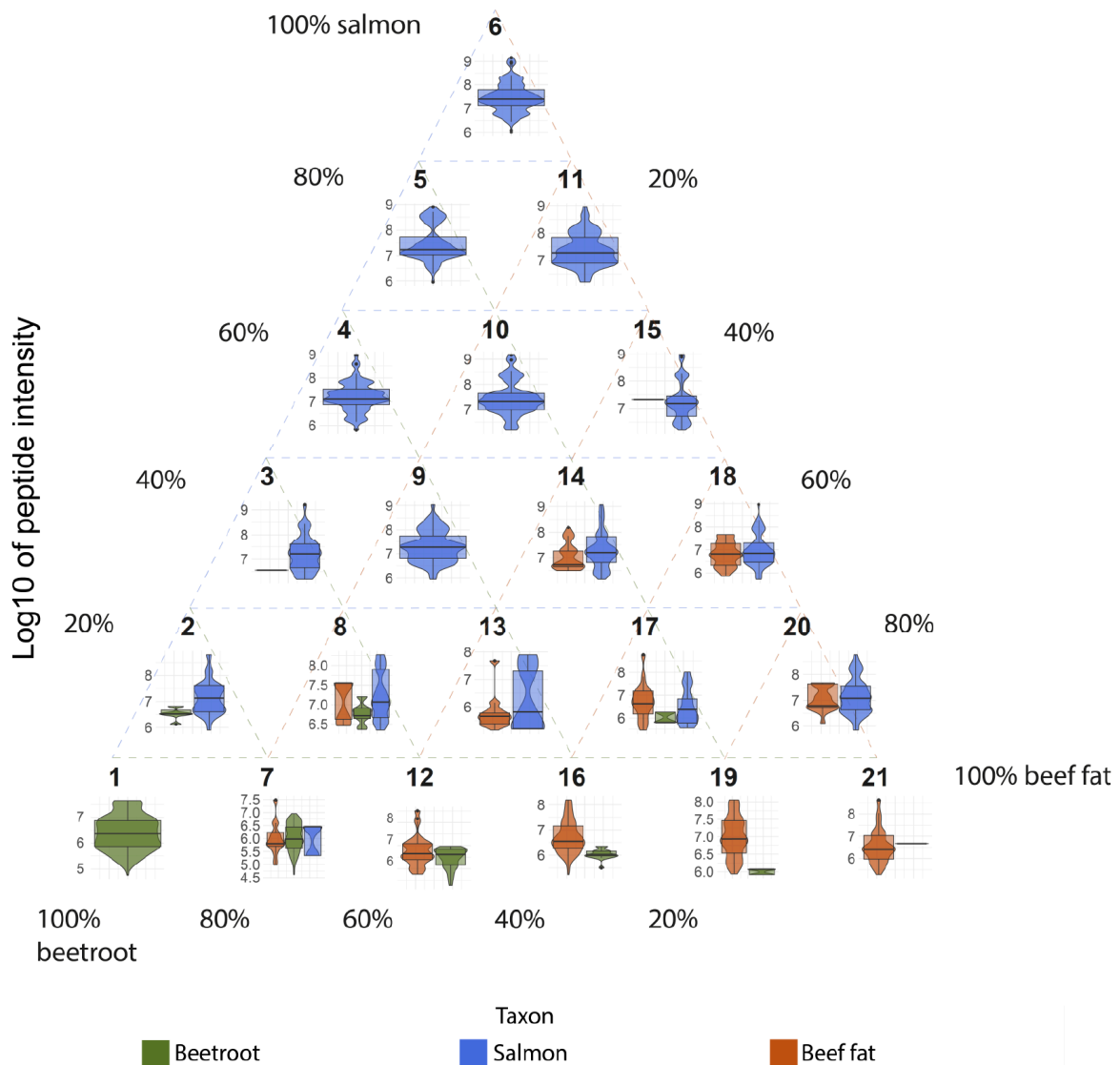


Figure 5, Distribution of the intensity of the three most intense peptides per protein. The y-axis shows the log 10 of intensity. For the ingredient combination of each experiment number see Fig. 1.

The quantification strategy described above is limited by the severe cooking- and burial-induced degradation of the proteins. The number of proteins that survive all cooking

events and the burial is limited across all conditions of mixed foods (Fig. 3). Out of the total 623 proteins that were identified across all the cooking experiments, 13 are present in samples from both the first, fifth, ninth cooking event and post-burial (Table 2). Remarkably, of these 13 persistent proteins, three belong to plant species, while none of the *B. taurus* proteins were found in samples from each cooking event. An overview of all identified proteins and the number of samples they were recovered in per cooking event (SI 5) shows that there are three additional salmon proteins that were recovered from the post-excavation samples but were not observed in samples from earlier time points. Two, B5DGS2 and B5X142, were not observed in the sample taken after the fifth cooking event, while A0A1S3QIK3 was not detected after the ninth cooking event. Additionally, some proteins appear to occur consistently and frequently throughout the experiment until the burial, after which they disappear. A notable example is two collagen-like entries, A0A1S3Q7E3 and A0A1S3SM10, which are found in 9-11 samples after the first, fifth and ninth cooking event, but in none after burial.

Table 2, Uniprot accession number, full name and associated species of the proteins found in at least one rim sample from each sampling point. The values indicate the number of samples each protein was identified in per sampling point.

Accession number	Protein Full Name	Organism	1	5	9	PE
B5DG55	Alpha-1,4 glucan phosphorylase	Salmo salar	16	10	11	4
A0A0J8E469	rRNA N-glycosylase	Beta vulgaris	12	2	8	2
B5DFU7	AMP deaminase	Salmo salar	10	6	7	2
A0A1S3PV75	Glucose-6-phosphate isomerase	Salmo salar	12	8	10	1
A0A1S3MVJ3	ATP synthase subunit beta	Salmo salar	14	5	7	1
Q7ZZN0	Myosin regulatory light chain 2	Salmo salar	15	4	7	1
B5X4K4	L-lactate dehydrogenase	Salmo salar	14	8	6	1
A0A1S3KNQ3	2-iminobutanoate/2-iminopropanoate deaminase	Salmo salar	5	5	6	1
B5DFX8	Phosphoglycerate kinase	Salmo salar	6	3	4	1
B5DGZ1	ADP/ATP translocase	Salmo salar	10	2	4	1
A0A0J8B2U8	Uncharacterized protein	Beta vulgaris	9	1	3	1
A0A803LHN7	Alcohol dehydrogenase	Chenopodium quinoa	7	1	3	1
B5XDB2	D-dopachrome decarboxylase	Salmo salar	7	2	2	1

The above results highlight the challenges in linking the abundance of the identified proteins with the species used in the cooking, which was one of the prime aims of this study. Of secondary importance was to identify the tissue types based on the identified proteins. One theoretical advantage of proteins over other biomolecules is that they may reflect particular tissue-types. In order to achieve this we used STRING (version 12.0) (Snel et al. 2000; Szklarczyk et al. 2023) to create protein networks for each of the three species and downloaded the tissue expression analysis, which compares the genes associated with the identified proteins to a database (TISSUE) of gene expression profiles per tissue type (Palasca et al. 2018). The output of this analysis is a list of tissue types, gene names and various metrics regarding the robustness of the tissue association. We filtered this list of tissue associations to only include tissues with a false discovery rate below 0.01 and a minimum number of 35 supporting proteins. The remaining tissue-gene associations were linked back to the list of proteins identified per sample (Fig. 6). Unfortunately, the TISSUE database did not include any data for the observed plant proteins, so it was not possible to associate any particular tissues with the plant proteins.

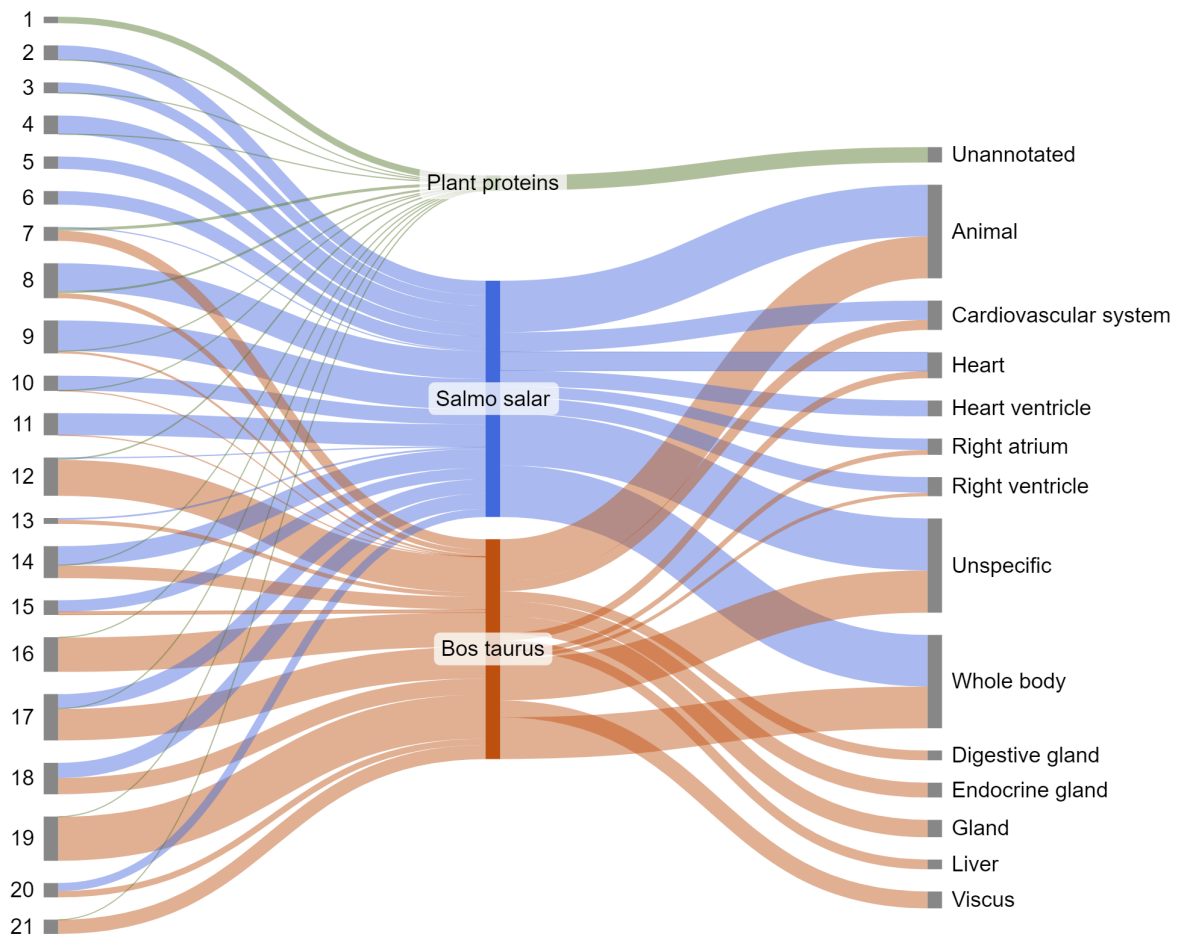


Figure 6, Tissue associations of the identified proteins per species.

The tissue associations from detected salmon and beef proteins are mostly unspecific, e.g. “animal” or “whole body”. Several associations with heart-related tissues, such as the heart itself, but also the right atrium and ventricles, were observed. One could argue to group these together as one large cluster associated with the cardiovascular system. We chose to keep them separate, as it highlights a disadvantage of this kind of tissue annotation, namely the unspecificity of the observed proteins. The fact that the average number rounded down of associated tissues per gene in our dataset is 5 ± 2 illustrates well how generic most of these genes are. This also explains why tissues such as the liver and various glands are listed in Fig. 6 as well. Despite the lack of a clear muscle-associated proteome, Fig. 6 highlights the difficulty in distinguishing between raw fats and muscle based on gene expression annotations, as well as showcases the breadth of tissue associations that can be expected from a foodcrust derived from these tissues.

Foodcrust accumulation showed sizeable variation between samples (Fig. 7). A linear regression model of the percentage of the ingredients and the foodcrust volume yielded no statistically significant relationships and had an adjusted R^2 value of 0.16, showing that the relative abundance of the three ingredients could only explain a small portion of the foodcrust volume variability. A linear regression model of the nutritional content showed a similar pattern. The overall model had a R^2 value of 0.16 and neither the absolute input of proteins, carbohydrates or fats yielded a significant relationship.

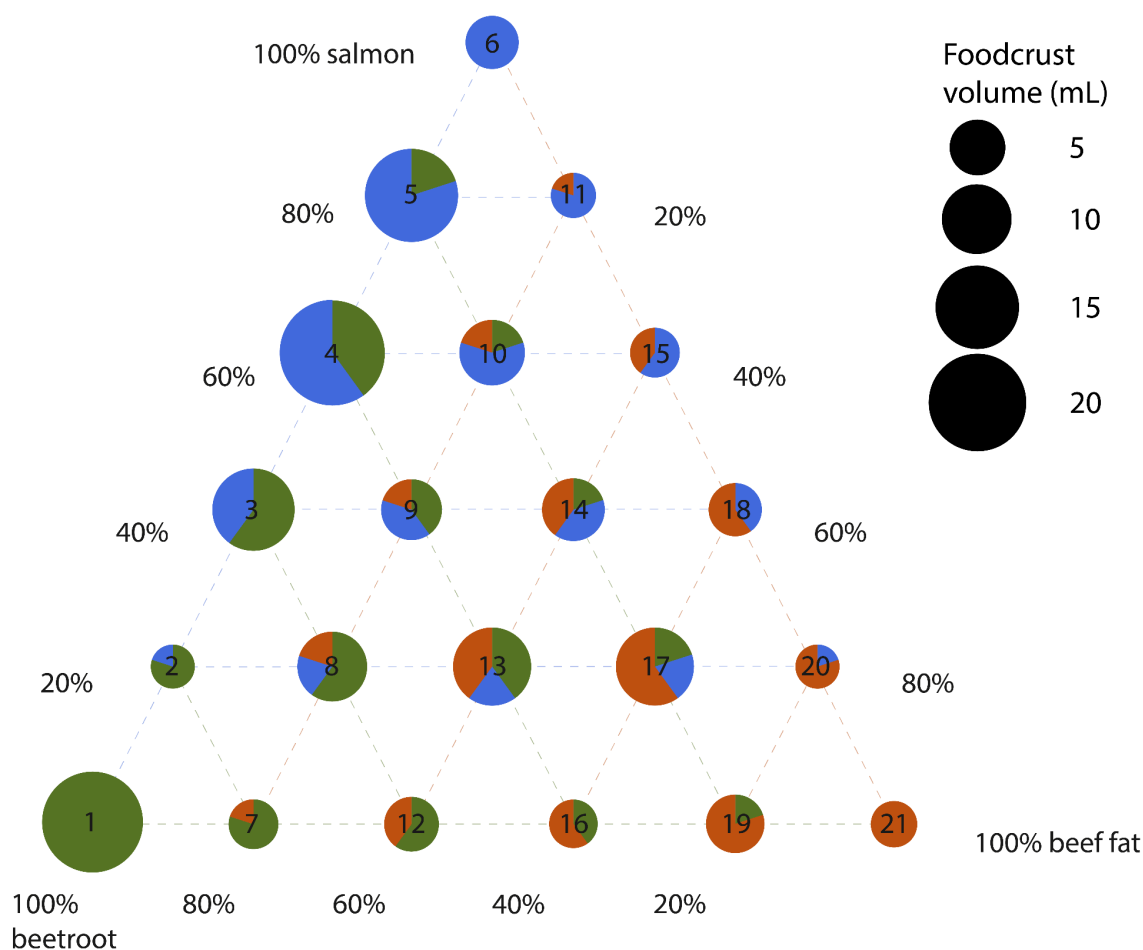


Figure 7, Volume of foodcrust accumulated after nine cooking events, sorted by experimental conditions listed in Fig. 1. The input ingredient composition of each of the experiments is shown by the pie chart on top of the bars. Increasing the proportion of one ingredient over another does not seem to have influenced foodcrust formation.

3.5 Discussion

3.5.1 Protein recovery

Our results show large variations in both the number of recovered proteins as well as the relative taxonomic composition of the recovered proteome between the different samples. However, the only significant relationship that we could observe was a decrease in the number of proteins after more cooking events. Such a decrease in protein recovery is not unexpected as it is well known that proteins are vulnerable to high temperatures (Ortner, vonEndt and Robinson 1972; Collins et al. 2002) and a decrease in protein recovery from fresh foodstuffs to foodcrust has been observed in previous studies (Evans et al. 2024). In fact, one of the aims of cooking is to make food more easily digestible by opening up closed protein structures (Bax et al. 2012; Yu et al. 2017). Unfortunately, denatured proteins are more vulnerable to degradation (Roberts et al. 2002; Zielbauer et al. 2016), which bodes ill for protein extraction from archaeological vessels. In the reference database we used in our analysis we did not include any soil bacteria, which are likely present in the samples post-excavation. The inclusion of bacteria in the reference database would likely boost the

number of proteins observed after excavation, considering that foodcrusts represent a rich source of nutrients for bacteria. Potential differences in the bacteria species that would colonise the foodcrusts depending on the cooked foods would have been interesting to investigate and potentially archaeologically relevant. However, it is unknown to what extent the soil bacteria at the YEAR centre can be compared to those from burial environments and it was beyond the scope of this study to establish a comprehensive and representative overview of soil microbiomes. Furthermore, the aim of this study was to test our ability to recover dietary proteins and to investigate the extent to which they represent the cooked foods. Including bacterial proteins in the analysis would have biased the number of recovered proteins towards the post-excavation samples, but would not have contributed to our understanding of how dietary proteins are represented in foodcrusts.

Although the decrease in the number of recovered proteins with increased cooking events demonstrates the negative effects of prolonged heating, the generally low number of proteins found after the first cooking event indicates that even limited exposure to heat can already induce severe protein degradation. Moreover, apart from the heating-induced damage, biases in the incorporation of proteins in the foodcrusts and extraction efficiency may have contributed to the limited size of the extracted proteome. The fact that the standard deviation is roughly equal to the mean number of detected proteins for cooking events 5, 9 and post-excavation, plus the large variation in how different experimental conditions respond to increased cooking in our opinion suggests that there might be considerable variation in protein preservation across the foodcrust in one pot and by only taking a sample from a single spot, our sampling strategy might have made our analysis vulnerable to this variation. The degree of spatial variation in foodcrust protein content might best be tested with a spectroscopy-based method, such as NIR (Maduro Dias et al. 2021). A better understanding of the amount of spatial variation as well as the scale at which it occurs could be of great benefit in maximising the proteomic data obtained from these degraded samples.

Furthermore, recent studies highlight that the current state of proteomic bioinformatic analysis only manages to assign a protein identification to a fraction of the extracted compounds (Chiang, Welker and Collins 2024; Palomo et al. 2023). It may be that future proteomic bioinformatic developments will significantly improve our ability to identify proteins in foodcrust samples, yet for now, we were only able to identify a small set of proteins, which for some samples do not represent all the taxa used in the cooking mixture.

All these issues may have impacted our ability to extract and identify proteins from our experimental samples, but ancient foodcrusts may have been exposed to higher temperatures, for longer periods of time and more often. Nevertheless, previous studies have shown that it is possible to extract proteins from archaeological foodcrusts, but the number of recovered proteins are often in the single digits (Shevchenko et al. 2018; Lyu et al. 2024). Considering the limitations of recovering proteins from foodcrusts, it is important to evaluate how well we can succeed in using the limited number of found proteins to identify the taxa and tissues used during the cooking process. Especially in comparison to lipid analysis, which is the most common method for analysing pottery use (Cubas et al. 2023).

3.5.2 Disentangling mixtures & taxonomic representativeness

One of the main theoretical advantages of proteomics over lipid analysis is the ability to distinguish the different resources in a mixture. As interpretations of resource use in lipid analysis are often based on compound-specific isotope ratios, samples with values that do

not fall in the established range of the references are more challenging to interpret. Lipid biomarkers could help with this problem, but they are relatively limited, both in terms of the number of taxa with established biomarkers as well as the taxonomic specificity of the biomarkers (Colonese et al. 2017; Bondetti et al. 2021; Whelton et al. 2021; Drieu et al. 2020). Alternatively, one can apply a mixing model to calculate what combination of foodstuffs could explain the observed values. Mixing models are known to work well in estimating the proportion of a limited number of resources with distinct isotope ratios but struggle more with more complex mixtures of a larger number of foodstuffs. Additionally, species with very similar ecological niches will be nearly impossible to distinguish isotopically (Cheung and Szpak 2021).

In theory, proteomic analysis does not suffer from these limitations. Mixing of multiple foodstuffs should not impact our ability to identify them and proteins can allow for up to species-level identifications (Rüther et al. 2022), although this is dependent on the type of recovered proteins and their coverage. Although these theoretical advantages of proteomics might lead to high expectations regarding the performance of proteomic analysis of foodcrust, the aforementioned issues in recovering proteins should temper them. To illustrate this point, 25 out of the 84 samples analysed in this study yielded no proteins at all. 32 samples yielded proteins from all species that were cooked in them. In 18 samples no proteins of one of the cooked species could be found and in 4 samples two species were missing. Lastly, in five samples proteins were found that belonged to a species that was not part of the mixture cooked in that pot (Table 1). The peptides that were uniquely associated with these proteins were run through BLAST (Altschul et al. 1990) to verify their taxonomic identification. The most specific taxon represented by at least two peptides was accepted as the new taxonomic identification (Table 3). This authentication step indicated that the plant proteins in samples 15.1 and 21.1 were shared up to the level of Eukaryota and could thus well derive from the ingredients cooked in the vessels. Although the BLAST taxonomic identifications are more specific for some of the remaining proteins than the others, it does appear that in samples 6.5, 7.1, 12.1 and 12.9 there are proteins present from a species that was not used in the cooking.

Table 3, Revised taxonomic identifications using BLAST of potentially intrusive proteins.

Experiment number	Cooking event	Protein Accession	Protein Full Name	Sequence Coverage	Number of Unique Peptides	Organism	BLAST taxon
6	5	A0A3Q1LL35	Uncharacterized protein	25%	2	Bos taurus	Boreoeutheria
7		A0A1S2X522	phosphopyruvate hydratase	63%	4	Salmo salar	Clupeocephala
	1	B5DG55	Alpha-1,4 glucan phosphorylase	48%	5	Salmo salar	Salmonidae
		Q3ZLR1	Superoxide dismutase [Cu-Zn]	23%	2	Salmo salar	Salmo salar
12	1	C0H9B8	Trifunctional enzyme subunit beta, mitochondrial	9%	2	Salmo salar	Teleostei
	9	A0A1S3Q7E3	collagen alpha-1(I) chain-like	10%	2	Salmo salar	Salmoninae
		A0A1S3SM10	collagen alpha-1(I) chain-like	8%	2	Salmo salar	Salmoninae
15	1	Q9LEE0	phosphopyruvate hydratase	15%	2	Spinacia oleracea	Eukaryota
21		A0A0J7YLM1	Elongation factor Tu (Fragment)	13%	4	Beta vulgaris subsp. vulgaris	Eukaryota
	1	A0A0J8B1E8	50S ribosomal protein L5, chloroplastic	14%	2	Beta vulgaris subsp. vulgaris	Eukaryota
		A0A0K9Q7T6	S1 motif domain-containing protein	4%	2	Spinacia oleracea	Eukaryota

There are several explanations for these ‘lost’ proteins, which occur in 5 out of 84 samples. It could be that the contamination occurred during the cooking experiment or during sampling, although sampling equipment was cleaned with ethanol after each sample. Contamination could have occurred during protein extraction, but all proteins present in the blank were not included in further analysis. Another potential explanation is machine carry over. The fish-specific peptides detected in samples 6.1, 12.1, 12.9 were all also detected in

the samples run before them. It may have happened that proteins from a previous sample remained stuck on the column and were only eluted with the next sample. Carry over seems a less likely explanation for the *B. taurus* protein in sample 6.5, as it was the first sample in its batch. Instead, it may be that this protein, A0A3Q1LL35, is background contamination. It is listed as an uncharacterised protein and the taxon assigned by BLAST, Boreoeutheria, encompasses most of the placental mammals. Consequently, this protein could have derived from any number of common contaminants that could have been introduced at any stage of the experiment.

As with the overall number of proteins, the representativeness of the extracted proteome decreases substantially with additional cooking episodes. After the first cooking event 5 of the 21 samples were missing one of the input foods, but after burying the samples 6 out of 7 samples were missing proteins from one or two of the cooked foods. Interestingly, despite 3 of 7 samples yielding protein after burial containing beef fat, none yielded *Bos taurus* proteins. To check if there was a systematic bias in recovery against any particular species we calculated the recovery rate of each species by dividing the number of samples they were found in by the number of samples that the species was cooked in (Table 4). The aforementioned 5 samples with proteins from species that were not part of the mixture were excluded from the recovery rate calculation for that particular species. The protein abundance of each species was not taken into account in calculating the recovery rate due to the previously discussed challenges in quantifying protein abundance for degraded samples.

Table 4, Recovery rate of each species per cooking event.

Cooking event	Beetroot	Salmon	Beef
1	80%	93%	87%
5	27%	93%	67%
9	53%	80%	73%
PE	13%	40%	0%
Average	43%	77%	57%

Table 4 shows clear differences in recovery rate between the foodstuffs. In general, beetroot was the most challenging to recover, which is in line with expectations considering the low protein content of beetroot. Vice versa, the fact that salmon has the highest recovery rate is in line with it having the highest protein content. Additionally, although in general there is a pattern of lower numbers of proteins and decreased recovery rate with an increased number of cooking events, the recovery rate for beetroot and beef is higher for cooking event 9 than 5. However, the Kruskal Wallis and Dunn's test showed that cooking events 5 and 9 were the only two groups without a statistically significant difference in protein numbers. The lack of a statistically significant difference may be explained by the addition of fresh food at the start of each cooking event. It may be that it is this 'freshest' food incorporated into the foodcrust that is yielding us proteins.

Secondly, Table 4 highlights the sudden disappearance of *Bos taurus* proteins after burial. It is unclear what causes the lack of *Bos taurus* protein preservation. It could be that the proteins in beef fat (which includes vimentin, creatine kinase B-type and heat shock protein beta-1) are simply more vulnerable to degradation due to their structure. Although Fig. 5 shows that the *Bos taurus* proteins identified are associated with largely the same

tissues as the *Salmo salar* proteins. Alternatively, the vulnerability to degradation may be due to the structure of the foodcrust and how fat-rich foods are incorporated in them rather than the structure of the proteins. There is some reason to suspect other factors might also be at play. A recent study performed a similar experiment, generating and burying foodcrusts made using red deer (*Cervus elaphus*), salmon (*Salmo salar*) and sweet chestnut (*Castanea sativa*) (Evans et al. 2024). Although our experiment did not use *Cervus elaphus*, it may be compared to our vessels with *Bos taurus*. The two species are closely enough related that most of their proteins should be similar, if not their sequences. Of their top 5 most abundant deer proteins in the foodcrust after cooking by peptide count, only fructose-bisphosphate aldolase (A0A3S5ZPB0) was detected in our samples and only 2 samples from the first cooking event. Two of the other common deer proteins, troponin T3 and myosin-1 were not directly detected in our samples, but myosin light chain 1/3 (A0JNJ5) and troponin I2 (F6QIC1) were detected, although again in a limited number of samples. As for the most abundant proteins after cooking, only myoglobin was detected in our samples, but bovine myoglobin is the 7th most frequently recovered bovine protein in our samples. However, it was not found in any samples after burial.

There is more overlap in the *Salmo salar* proteins between our study and Evans *et al.* (2024). All five most abundant *Salmo salar* proteins from their after-cooking samples were observed in our samples, although they do not occur in the top 10 of most frequently recovered *Salmo salar* proteins in our samples. As for the most abundant proteins after burial, four out of five were detected in our samples, but only one, myosin heavy chain, fast skeletal muscle-like (A0A1S3QIK3), in one of our after-burial samples.

The recovery of sweet chestnut and beetroot cannot be compared as easily as *Cervus elaphus* and *Bos taurus*, but it is noteworthy that they recover chestnut proteins less frequently than *Salmo salar* or *Cervus elaphus* proteins in their buried samples, whereas we do find plant proteins, but no *Bos taurus* proteins in our buried samples. Of course, their experiment used deer meat, while we used beef fat and it may be that this changes the preservation potential of the proteins contained within the tissue. Regardless of what prevented us from recovering *Bos taurus* proteins, it is clear that, especially for buried samples, the recovered proteome might not reflect all the foodstuffs cooked in the vessel and that there is substantial variability in our ability to obtain proteins from buried samples.

3.5.3 Foodcrust abundance

Our and previous results (Evans et al. 2024) show that the protein contents of foodcrusts do not ‘fairly’ represent all the foodstuffs prepared in the vessel. Biases in protein recovery for particular foodstuffs were expected to a certain extent, but Fig. 6 suggests that there are also differences in the quantity of foodcrust that particular foodstuffs generate. Fig. 6 shows that there is not one particular food that is driving the formation of more foodcrust. Overall, the beetroot-salmon mixtures produce the most foodcrust, followed by beetroot-salmon-beef fat mixtures. Beetroot-beef fat or salmon-beef fat seems to produce the least foodcrust, suggesting that the presence of such a fat-rich food is not necessarily a driver of foodcrust formation.

The key finding of Fig. 5 is perhaps to highlight the importance of carbohydrate and protein-rich foodstuffs in generating foodcrust in the first place, while fat-rich foods seem to play a lesser role. Yet the most common method of analysing these organic residues has so far been lipid analysis, which may bias our interpretation of the processed resources towards fat-rich foodstuffs. Such a variable pattern in foodcrust formation in our dataset emphasises

the value of replicates, which we, unfortunately, are not able to provide. Nonetheless, we argue that these results are a valuable addition to our current understanding of the formation of organic residues, especially because as far as we are aware there has been no absolute quantification of the abundance of foodcrust formed up to this date.

3.6 Conclusion

This experiment had three core aims: firstly, to determine the relationship between the relative ratios of input foods and the proportions between the species of the recovered proteins. Secondly, to measure the impact of repeated cooking and burial on protein preservation and lastly how the different ingredient combinations impacted the amount of foodcrust that formed.

The data shows clearly that the relative abundance of recovered proteins does not mirror the actual input of foodstuffs. Most identified proteins belonged to *Salmo salar*, followed by *Bos taurus* and lastly *Beta vulgaris*. This pattern applied to samples taken after cooking events 1, 5 and 9. However, in the samples taken after the burial no *Bos taurus* proteins could be recovered, although a small number of *Beta vulgaris* proteins were identifiable. These results highlight that not only do food input and recovered proteins not necessarily align, the relative abundance of taxa may change due to increased degradation.

Secondly, the number of recovered proteins decreased significantly from the first to the fifth cooking event, plateaued from the fifth to the ninth and decreased again from the ninth event until after burial.

Lastly, no statistically significant predictor of the foodcrust volume could be identified, but the mixtures that formed the most foodcrust in terms of volume were beetroot-salmon, followed by a mixture of all three ingredients, suggesting that lipid-rich foods are not as important for the volume of foodcrust.

References

- Admiraal, M. et al. (2020). Leftovers: The presence of manufacture-derived aquatic lipids in Alaskan pottery. *Archaeometry*, 62 (2), Wiley., pp.346–361. [Online]. Available at: doi:10.1111/arcm.12515.
- Admiraal, M. et al. (2023). Chemical analysis of pottery reveals the transition from a maritime to a plant-based economy in pre-colonial coastal Brazil. *Scientific reports*, 13 (1), p.16771. [Online]. Available at: doi:10.1038/s41598-023-42662-5.
- Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410. [Online]. Available at: doi:10.1016/S0022-2836(05)80360-2.
- Astrup, P. M. et al. (2021). A drowned Mesolithic shell midden complex at Hjørnø Vesterhoved, Denmark and its wider significance. *Quaternary science reviews*, 258, p.106854. [Online]. Available at: doi:10.1016/j.quascirev.2021.106854.
- Bax, M.-L. et al. (2012). Cooking temperature is a key determinant of in vitro meat protein digestion rate: investigation of underlying mechanisms. *Journal of agricultural and food chemistry*, 60 (10), pp.2569–2576. [Online]. Available at: doi:10.1021/jf205280y.
- Bleasdale, M. et al. (2021). Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nature communications*, 12 (1), nature.com., p.632. [Online]. Available at: doi:10.1038/s41467-020-20682-3.
- Bondetti, M. et al. (2020). Resource processing, early pottery and the emergence of Kitoi culture in Cis-Baikal: Insights from lipid residue analysis of an Early Neolithic ceramic assemblage from the Gorelyi Les habitation site, Eastern Siberia. *Archaeological research in Asia*, 24 (100225), Elsevier BV., p.100225. [Online]. Available at: doi:10.1016/j.ara.2020.100225.
- Bondetti, M. et al. (2021). Investigating the formation and diagnostic value of ω -(o-alkylphenyl)alkanoic acids in ancient pottery. *Archaeometry*, 63 (3), pp.594–608. [Online]. Available at: doi:10.1111/arcm.12631.
- Charlton, S. et al. (2019). New insights into Neolithic milk consumption through proteomic analysis of dental calculus. *Archaeological and anthropological sciences*, 11 (11), Springer., pp.6183–6196. [Online]. Available at: doi:10.1007/s12520-019-00911-7.
- Cheung, C. and Szpak, P. (2021). Interpreting past human diets using stable isotope mixing models. *Journal of archaeological method and theory*, 28 (4), Springer Science and Business Media LLC., pp.1106–1142. [Online]. Available at: doi:10.1007/s10816-020-09492-5.
- Chiang, Y., Welker, F. and Collins, M. J. (2024). Spectra without stories: reporting 94% dark and unidentified ancient proteomes. *Open research Europe*, 4, p.71. [Online]. Available at: doi:10.12688/openreseurope.17225.1.
- Cleland, T. P. (2018). Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *Journal of proteome research*, 17 (11), pp.3976–3983. [Online]. Available at: doi:10.1021/acs.jproteome.8b00637.
- Collins, M. J. et al. (2002). The survival of organic matter in bone: a review. *Archaeometry*, 44 (3), Wiley., pp.383–394. [Online]. Available at: doi:10.1111/1475-4754.t01-1-00071.

Colonese, A. C. et al. (2017). New criteria for the molecular identification of cereal grains associated with archaeological artefacts. *Scientific reports*, 7 (1), p.6633. [Online]. Available at: doi:10.1038/s41598-017-06390-x.

Copley, M. S. et al. (2003). Direct chemical evidence for widespread dairying in prehistoric Britain. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (4), Proceedings of the National Academy of Sciences., pp.1524–1529. [Online]. Available at: doi:10.1073/pnas.0335955100.

Copley, M. S. et al. (2005). Processing of milk products in pottery vessels through British prehistory. *Antiquity*, 79 (306), Cambridge University Press., pp.895–908. [Online]. Available at: doi:10.1017/S0003598X00115029 [Accessed 12 May 2022].

Courel, B. et al. (2020). Organic residue analysis shows sub-regional patterns in the use of pottery by Northern European hunter–gatherers. *Royal Society Open Science*, 7 (4), Royal Society., p.192016. [Online]. Available at: doi:10.1098/rsos.192016.

Courel, B. et al. (2021). The use of early pottery by hunter-gatherers of the Eastern European forest-steppe. *Quaternary science reviews*, 269, p.107143. [Online]. Available at: doi:10.1016/j.quascirev.2021.107143.

Craig, O. E. et al. (2005). Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity*, cambridge.org. [Online]. Available at: <https://www.cambridge.org/core/journals/antiquity/article/did-the-first-farmers-of-central-and-eastern-europe-produce-dairy-foods/284138196CFD83FA06340C061EDF5F93>.

Craig, O. E. et al. (2011). Ancient lipids reveal continuity in culinary practices across the transition to agriculture in Northern Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (44), pp.17910–17915. [Online]. Available at: doi:10.1073/pnas.1107202108.

Craig, O. E. et al. (2013). Earliest evidence for the use of pottery. *Nature*, Nature Publishing Group. [Online]. Available at: doi:10.1038/nature12109.

Cronin, T. et al. (2007). Composition of ancient Irish bog butter. *International dairy journal*, 17 (9), Elsevier BV., pp.1011–1020. [Online]. Available at: doi:10.1016/j.idairyj.2007.01.012.

Cubas, M. et al. (2023). Organic Residue Analysis of Archaeological Pottery. In: *Reference Module in Social Sciences*. Elsevier. [Online]. Available at: doi:10.1016/B978-0-323-90799-6.00047-1.

Dekker, J. et al. (2023). Spatial analysis of the ancient proteome of archeological teeth using mass spectrometry imaging. *Rapid communications in mass spectrometry: RCM*, 37 (8), p.e9486. [Online]. Available at: doi:10.1002/rcm.9486.

Dinno, A. (2017). *dunn.test: Dunn's Test in dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums*. [Online]. Available at: <https://rdrr.io/cran/dunn.test/man/dunn.test.html> [Accessed 5 March 2024].

Drieu, L. et al. (2020). Is it possible to identify ancient wine production using biomolecular approaches? *STAR: Science & Technology of Archaeological Research*, 6 (1), Routledge., pp.16–29. [Online]. Available at: doi:10.1080/20548923.2020.1738728.

Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association*, 56 (293), Taylor & Francis., pp.52–64. [Online]. Available at: doi:10.1080/01621459.1961.10482090.

- Enghoff, I. (2011). *Regionality and biotope exploitation in Danish Ertebølle and adjoining periods*, Scientia Danica. Series B Biologica. Det Kongelige Danske Videnskabernes Selskab.
- Evans, M. et al. (2023). Detection of dairy products from multiple taxa in Late Neolithic pottery from Poland: an integrated biomolecular approach. *Royal Society open science*, 10 (3), p.230124. [Online]. Available at: doi:10.1098/rsos.230124.
- Evans, M. et al. (2024). The impact of cooking and burial on proteins: a characterisation of experimental foodcrusts and ceramics. *bioRxiv*. [Online]. Available at: doi:10.1101/2024.04.03.587902.
- Greco, E. et al. (2018). Proteomic Analyses on an Ancient Egyptian Cheese and Biomolecular Evidence of Brucellosis. *Analytical chemistry*, 90 (16), pp.9673–9676. [Online]. Available at: doi:10.1021/acs.analchem.8b02535.
- Harvey, V. L. and Hendy, J. (2025). *SP3 protocol optimised for foodcrust protein extraction from archaeological cooking vessels*. [Online]. Available at: <https://www.protocols.io/view/sp3-protocol-optimised-for-foodcrust-protein-extra-dyge7tte.pdf> [Accessed 31 January 2025].
- Hendy, J. et al. (2018a). A guide to ancient protein studies. *Nature ecology & evolution*, 2 (5), nature.com., pp.791–799. [Online]. Available at: doi:10.1038/s41559-018-0510-x.
- Hendy, J. et al. (2018b). Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nature communications*, 9 (1), p.4064. [Online]. Available at: doi:10.1038/s41467-018-06335-6.
- Holst, D. (2010). Hazelnut economy of early Holocene hunter–gatherers: a case study from Mesolithic Duvensee, northern Germany. *Journal of archaeological science*, 37 (11), Elsevier BV., pp.2871–2880. [Online]. Available at: doi:10.1016/j.jas.2010.06.028 [Accessed 23 July 2024].
- Kruskal, W. H. and Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47 (260), Taylor & Francis., pp.583–621. [Online]. Available at: doi:10.1080/01621459.1952.10483441.
- Kubiak-Martens, L. (1999). The plant food component of the diet at the late Mesolithic (Ertebølle) settlement at Tybrind Vig, Denmark. *Vegetation history and archaeobotany*, 8 (1-2), Springer Science and Business Media LLC., pp.117–127. [Online]. Available at: doi:10.1007/bf02042850.
- Kubiak-Martens, L. (2002). New evidence for the use of root foods in pre-agrarian subsistence recovered from the late Mesolithic site at Halsskov, Denmark. *Vegetation history and archaeobotany*, 11 (1-2), Springer Science and Business Media LLC., pp.23–32. [Online]. Available at: doi:10.1007/s003340200003.
- Kubiak-Martens, L. et al. (2024). Transformation of cereal grains: Botanical and chemical analysis of food residues encrusted on pottery from the Funnel Beaker settlement of Oldenburg LA 77, northern Germany. *PloS one*, 19 (1), p.e0296986. [Online]. Available at: doi:10.1371/journal.pone.0296986.
- Kubiak-Martens, L., Brinkkemper, O. and Oudemans, T. F. M. (2015). What's for dinner? Processed food in the coastal area of the northern Netherlands in the Late Neolithic. *Vegetation history and archaeobotany*, 24 (1), Springer Science and Business Media LLC., pp.47–62. [Online]. Available at: doi:10.1007/s00334-014-0485-8.

Lucquin, A. et al. (2023). The impact of farming on prehistoric culinary practices throughout Northern Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 120 (43), p.e2310138120. [Online]. Available at: doi:10.1073/pnas.2310138120.

Lundy, J. et al. (2023). Cuisine in transition? Organic residue analysis of domestic containers from 9th-14th century Sicily. *Royal Society open science*, 10 (3), p.221305. [Online]. Available at: doi:10.1098/rsos.221305.

Lyu, N. et al. (2024). Characterization of Pottery Foodcrusts Through Lipid and Proteomic Analyses: A Case Study from the Xiawan Site in Yixing City, East China. *Journal of archaeological science*, 161 (105902), papers.ssrn.com. [Online]. Available at: doi:10.1016/j.jas.2023.105902 [Accessed 8 September 2023].

Maduro Dias, C. S. A. M. et al. (2021). Application of Near Infrared Reflectance (NIR) spectroscopy to predict the moisture, protein, and fat content of beef for gourmet hamburger preparation. *Livestock science*, 254 (104772), Elsevier BV., p.104772. [Online]. Available at: doi:10.1016/j.livsci.2021.104772 [Accessed 26 October 2024].

Mason, S. (2004). Archaeobotanical Analysis-Møllegabet II. In: Skaarup, J. and Grøn (Eds). *Møllegabet II, A submerged Mesolithic settlement in southern Denmark*. BAR International Series. 1328. Archaeopress. pp.122–143. [Online]. Available at: https://www.researchgate.net/profile/Ole-Gron/publication/267639126_Mollegabet_II_A_submerged_Mesolithic_settlement_in_southern_Denmark_With_contributions_by_Sarah_Mason_Lisa_Hodgetts_Peter_Rowley-Conwy_and_Annica_Cardell/links/54cb50e70cf22f98631e72fc/Mollegabet-II-A-submerged-Mesolithic-settlement-in-southern-Denmark-With-contribution-s-by-Sarah-Mason-Lisa-Hodgetts-Peter-Rowley-Conwy-and-Annica-Cardell.pdf#page=131.

Oras, E. et al. (2017). The adoption of pottery by north-east European hunter-gatherers: Evidence from lipid residue analysis. *Journal of archaeological science*, 78, pp.112–119. [Online]. Available at: doi:10.1016/j.jas.2016.11.010.

Ortner, D. J., vonEndt, D. W. and Robinson, M. S. (1972). The Effect of Temperature on Protein Decay in Bone: Its Significance in Nitrogen Dating of Archaeological Specimens. *American antiquity*, 37 (4), Cambridge University Press., pp.514–520. [Online]. Available at: doi:10.2307/278957 [Accessed 9 July 2024].

Page, L. M. and Burr, B. M. (2011). *Peterson Field Guide to Freshwater Fishes of North America North of Mexico*. Houghton Mifflin Harcourt. [Online]. Available at: <https://play.google.com/store/books/details?id=rFTqmdDqn-wC>.

Palasca, O. et al. (2018). TISSUES 2.0: an integrative web resource on mammalian tissue expression. *Database: the journal of biological databases and curation*, 2018. [Online]. Available at: doi:10.1093/database/bay003.

Pal Chowdhury, M., Campbell, S. and Buckley, M. (2021). Proteomic analysis of archaeological ceramics from Tell Khaiber, southern Iraq. *Journal of archaeological science*, 132, p.105414. [Online]. Available at: doi:10.1016/j.jas.2021.105414.

Palomo, I. R. et al. (2023). Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *bioRxiv*, p.2023.12.15.571577. [Online]. Available at: doi:10.1101/2023.12.15.571577 [Accessed 15 January 2024].

Posit team. (2024). *RStudio: Integrated Development Environment for R*. Boston, MA : Posit Software, PBC. [Online]. Available at: <http://www.posit.co/>.

Price, D. T. et al. (2001). Smakkerup Huse: A Mesolithic Settlement in NW Zealand,

Denmark. *Journal of Field Archaeology*, 28 (1-2), Routledge., pp.45–67. [Online]. Available at: doi:10.1179/jfa.2001.28.1-2.45.

Price, T. D. et al. (2018). Asnæs Havnepark: a late Mesolithic Ertebølle coastal site in western Sjælland, Denmark. *Danish Journal of Archaeology*, 7 (2), Routledge., pp.255–276. [Online]. Available at: doi:10.1080/21662282.2018.1551979.

Public Health England. (2021). *McCance and Widdowson's The Composition of Foods Integrated Dataset 2021*. [Online]. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/971021/McCance_and_Widdowsons_Composition_of_Foods_integrated_dataset_2021.pdf [Accessed 28 March 2022].

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing. [Online]. Available at: <https://www.R-project.org/>.

Richter, J. and Noe-Nygaard, N. (2003). A Late Mesolithic Hunting Station at Agernæs, Fyn, Denmark: Differentiation and Specialization in the late Ertebølle-Culture, heralding the Introduction of Agriculture? *Acta Archaeologica*, 74 (1), Brill., pp.1–32. [Online]. Available at: doi:10.1111/j.0065-001X.2003.aar740101.x [Accessed 13 December 2022].

Roberts, S. J. et al. (2002). The taphonomy of cooked bone: characterizing boiling and its physico-chemical effects. *Archaeometry*, 44 (3), Wiley., pp.485–494. [Online]. Available at: doi:10.1111/1475-4754.t01-1-00080.

Robinson, D. E. and Harild, J. A. (2009). Archaeobotany of an early Ertebølle (late Mesolithic) site at Halsskov, Zealand, Denmark. In: Mason, S. L. R. and Hather, J. (Eds). *Hunter-Gatherer Archaeobotany: Perspectives from the Northern Temperate Zone*. London, England : Routledge. [Online]. Available at: <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315427171-8/archaeobotany-early-erteb%C3%B8lle-late-mesolithic-site-halsskov-zealand-denmark-david-earle-robinson-jan-andreas-harild>.

Rüther, P. L. et al. (2022). SPIN enables high throughput species identification of archaeological bone by proteomics. *Nature communications*, 13 (1), nature.com., p.2458. [Online]. Available at: doi:10.1038/s41467-022-30097-x.

Schroeter, E. R. et al. (2016). Bone protein ‘extractomics’: comparing the efficiency of bone protein extractions of Gallus gallus in tandem mass spectrometry, with an eye towards paleoproteomics. *PeerJ*, 4, PeerJ Inc., p.e2603. [Online]. Available at: doi:10.7717/peerj.2603 [Accessed 5 April 2022].

Shevchenko, A. et al. (2018). Fine Endmesolithic fish caviar meal discovered by proteomics in foodcrusts from archaeological site Friesack 4 (Brandenburg, Germany). *PloS one*, 13 (11), p.e0206483. [Online]. Available at: doi:10.1371/journal.pone.0206483.

Snel, B. et al. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic acids research*, 28 (18), pp.3442–3444. [Online]. Available at: doi:10.1093/nar/28.18.3442.

Solazzo, C. et al. (2008). Identification of protein remains in archaeological potsherds by proteomics. *Analytical chemistry*, 80 (12), pp.4590–4597. [Online]. Available at: doi:10.1021/ac800515v.

Solntsev, S. K. et al. (2018). Enhanced Global Post-translational Modification Discovery with

MetaMorpheus. *Journal of proteome research*, 17 (5), ACS Publications., pp.1844–1851. [Online]. Available at: doi:10.1021/acs.jproteome.7b00873.

Szklarczyk, D. et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, 51 (D1), pp.D638–D646. [Online]. Available at: doi:10.1093/nar/gkac1000.

Tanasi, D. et al. (2021). Paleoproteomic profiling of organic residues on prehistoric pottery from Malta. *Amino acids*, 53 (2), pp.295–312. [Online]. Available at: doi:10.1007/s00726-021-02946-4.

The Global Proteome Machine Organization. (2012). *cRAP protein sequences*. [Online]. Available at: <https://www.thegpm.org/crap/> [Accessed 15 February 2024].

UniProt Consortium. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic acids research*, 51 (D1), pp.D523–D531. [Online]. Available at: doi:10.1093/nar/gkac1052.

Warinner, C. et al. (2014). Direct evidence of milk consumption from ancient human dental calculus. *Scientific reports*, 4, nature.com., p.7104. [Online]. Available at: doi:10.1038/srep07104.

Whelton, H. L. et al. (2021). A call for caution in the analysis of lipids and other small biomolecules from archaeological contexts. *Journal of archaeological science*, 132, Elsevier., p.105397. [Online]. Available at: doi:10.1016/j.jas.2021.105397.

Yang, Y. et al. (2014). Proteomics evidence for kefir dairy in Early Bronze Age China. *Journal of archaeological science*, 45, pp.178–186. [Online]. Available at: doi:10.1016/j.jas.2014.02.005.

Yu, T.-Y. et al. (2017). Cooking-induced protein modifications in meat. *Comprehensive Reviews in Food Science and Food Safety*, 16 (1), Wiley., pp.141–159. [Online]. Available at: doi:10.1111/1541-4337.12243.

Zielbauer, B. I. et al. (2016). Physical aspects of meat cooking: Time dependent thermal protein denaturation and water loss. *Food biophysics*, 11 (1), Springer Science and Business Media LLC., pp.34–42. [Online]. Available at: doi:10.1007/s11483-015-9410-7.

Harvey, V. and J. Hendy (2022). Unpublished protein extraction protocol shared by Virginia Harvey, 22 January 2022.

4 Trial by fire, a blind test of the proteomic analysis of experimental foodcrust

Joannes Adrianus Antonius Dekker^{1,2}, Max Ramsøe², Yun Chiang^{2,3}, Benjamin Dartigues⁴, Alexis Groppi⁴, Bente Philippsen⁵, John Meadows^{6,7}, Chloë Baldreki⁸, Tina Ravnsborg⁹, Ole Jensen⁹, Matthew Collins^{2,10}, Jessica Hendy¹

1. BioArch, University of York, UK.
2. Section for Geobiology, Globe Institute, University of Copenhagen, Denmark.
3. The Nice Institute of Chemistry, Universite Cote d'Azur, France.
4. Bordeaux Bioinformatics Center, University of Bordeaux, France.
5. National Laboratory for Age Determination, Norwegian University of Science and Technology, Norway.
6. LEIZA-ZBSA Zentrum für Baltische und Skandinavische Archäologie, Germany.
7. Christian Albrechts University Leibniz-Laboratory for AMS Dating and Stable Isotope Research, Germany.
8. Department of Biology, Bioscience Technology Facility, University of York, UK.
9. Institut for Biokemi og Molekylær Biologi, University of Southern Denmark, Denmark.
10. McDonald Institute for Archaeological Research, University of Cambridge, UK.

Written with intent of submission to PLOS ONE. Intended date of submission is 28-02-2025.

CRedit statement

J.D: Conceptualisation, Methodology, Validation, Formal analysis, Investigation, Investigation, Writing - Original draft, Writing - Review & editing, Visualisation, Project administration. **M.R:** Methodology, Investigation, Resources. **Y.C:** Software, Validation. **B.D:** Software, Validation, Formal analysis, Resources. **A.G:** Supervision, Project administration. **B.P:** Resources, Data curation. **J.M:** Resources, Data curation. **C.B:** Validation, Investigation, Resources, Writing - Original draft. **T.R:** Validation, Investigation, Resources, Writing - Original draft. **O.J:** Resources, Supervision, Project administration. **M.C:** Resources, Writing - Review & editing, Supervision, Project administration, Funding acquisition. **J.H:** Resources, Writing - Review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements:

We would like to thank the Bioscience Technology Facility and Chemistry Department at the University of York for mass spectrometry access and support. The instrumentation is part of the York Centre of Excellence in Mass Spectrometry. The centre was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1).

Additionally, we would like to thank Dr. Mansji Pal Chowdhury for his advice and helpful comments on how to apply the DES protocol.

Funding statement:

J.D. is funded by a Marie Skłodowska-Curie grant (grant agreement No. 956351) from the EU Horizon 2020 programme.

Abstract

Foodcrusts are a valuable source of information regarding past culinary practices. They can provide information about both the animal and the plant component of the diet and provide direct evidence of heat processing of the resources they contain. These reasons make foodcrusts an interesting target for palaeoproteomic analysis, as proteomic techniques are expected to be able to distinguish the different components of a food mixture and have a higher taxonomic resolution than other commonly used methods of organic residue analysis. However, in contrast to the more commonly analysed biomineral tissues, the proteins in foodcrust are not protected by a mineral matrix and are likely to have sustained damage from the heating process. Therefore, it is not clear whether proteins from foodcrust can live up to their expectations.

To test the capabilities of proteomics to accurately identify foodstuffs from foodcrust, we applied several proteomic workflows, including two extraction protocols (SP3 and DES), different mass spectrometers (timsTOF and Orbitrap) and bioinformatic validation software to analyse 20 experimental foodcrusts. Our results emphasise the great difficulty of accurately identifying the resources used to generate foodcrusts regardless of the used workflow. Following standard identification criteria the correct accurate identifications could only be made for three out of the twenty samples. In all other samples either certain foodstuffs were missing or false positives were identified. The vast majority of false positives could be removed by discarding all proteins with less than six associated peptides and by summarising all taxonomic identifications to order level, but this greatly reduces the informative value of the proteomic analysis.

4.1 Introduction

Ceramic fragments are one of the most commonly found and frequently studied types of archaeological artefacts. Obtaining direct evidence for what ceramic vessels were used for, however, remains challenging. Visual methods such as use-wear analysis are able to glean some information from microscopic traces left on the ceramics, though it is more suitable to identify the manner of use and cooking rather than the food that was cooked (Forte, Nunziante Cesaro and Medeghini 2018; Van Gijn et al. 2020). On the other hand biomolecular techniques are able to identify foodstuffs in ancient pottery and have proven essential in revolutionising our understanding of the use of ceramics in the past (Evershed et al. 2008; Craig et al. 2013, 2011; Salque et al. 2013). The most commonly used biomolecular technique is lipid analysis, which can be applied to lipid residues either trapped within the ceramic matrix or charred food residues (foodcrusts) adhering to the ceramic wall. Lipid residue analysis is a well-established and robust technique with many successful applications across the globe and on material from a wide variety of ages (Copley et al. 2004; Lucquin et al. 2016; Admiraal et al. 2020; Montalvo-Cabrera et al. 2024; Lundy et al. 2024). Its limitations lie primarily in the specificity of the resources it can distinguish. For the $\delta^{13}\text{C}$ ratio of the C16 and C18 alkanes, frequently used in lipid analysis, there are reference ranges for marine, freshwater, ruminant dairy, ruminant adipose, porcine fats, C3 and C4

plants, although some of these ranges overlap. It is therefore often possible to distinguish between these broad categories, but not to specify the taxon responsible for the lipid signal. Lipid biomarkers can be more specific, but these are only available for a relatively select number of species (Colonese et al. 2017; Drieu et al. 2020; Bondetti et al. 2021), even though current research strives to increase this number.

Increased taxonomic resolution in identifying the contents of cooking pots would be of great benefit to the study of past diets and proteins have been suggested as a suitable target compound to achieve this. The extraction of proteins from organic residues has a much shorter development history than lipids (Craig and Collins 2002; Solazzo et al. 2008; Barker et al. 2012) and although more and more successful applications are being published (Shevchenko et al. 2018; Tanasi et al. 2021; Evans et al. 2023; Lyu et al. 2024), the method is still in its infancy in some ways. For example, whereas lipid analysis routinely manages to successfully extract lipids from the ceramic matrix and foodcrusts, the extraction of proteins from ceramic matrix is extremely challenging and rarely successful, although some cases have been published (Chowdhury, Campbell and Buckley 2021; Chowdhury et al. 2022; Hendy et al. 2018b). Consequently, most proteomic studies focus either on calcified deposits (Hendy et al. 2018b; Evans et al. 2023) or foodcrusts (Shevchenko et al. 2018; Lyu et al. 2024).

Although the extraction of proteins from foodcrusts and other matrices is an important issue, perhaps a more pressing question is to what extent the extracted proteins reflect the cooked contents of a ceramic vessel. In particular, is proteomic analysis able to separate the individual tissue types and taxa from food mixtures? And lastly, if proteomics are able to correctly identify the different foodstuffs, to what taxonomic level can the identification be made? The aforementioned questions have all been reported as advantages of proteomics over lipid analysis (Hendy et al. 2018b; Cubas et al. 2023), but they have not yet been put to the test. Proteomic cooking experiments have been performed, but these often focused on protein recovery from single-ingredient remains (Dallongeville et al. 2011; Evans et al. 2024) and restricted their reference protein databases to only the species they knew were present in the foodcrusts (as was done in Chapter 3).

As proteomics are becoming an increasingly popular method to study diet using either proteins preserved in foodcrust, calcified deposits or the ceramic matrix itself, it is essential to establish the limitations of the method. To that end, we have performed a blind test on 20 modern foodcrusts generated in cooking experiments testing the ability of proteomics to identify foodstuffs in two sets of experimental foodcrusts. We also test two different extraction protocols, SP3 and DES, as well as two types of mass spectrometers, orbitrap and timsTOF and a number of software tools (Fig. 1). The aim of using this variety of tools and protocols is not to crown a single workflow as the “best”, but rather to investigate the degree of consistency in the identification of foodstuffs.

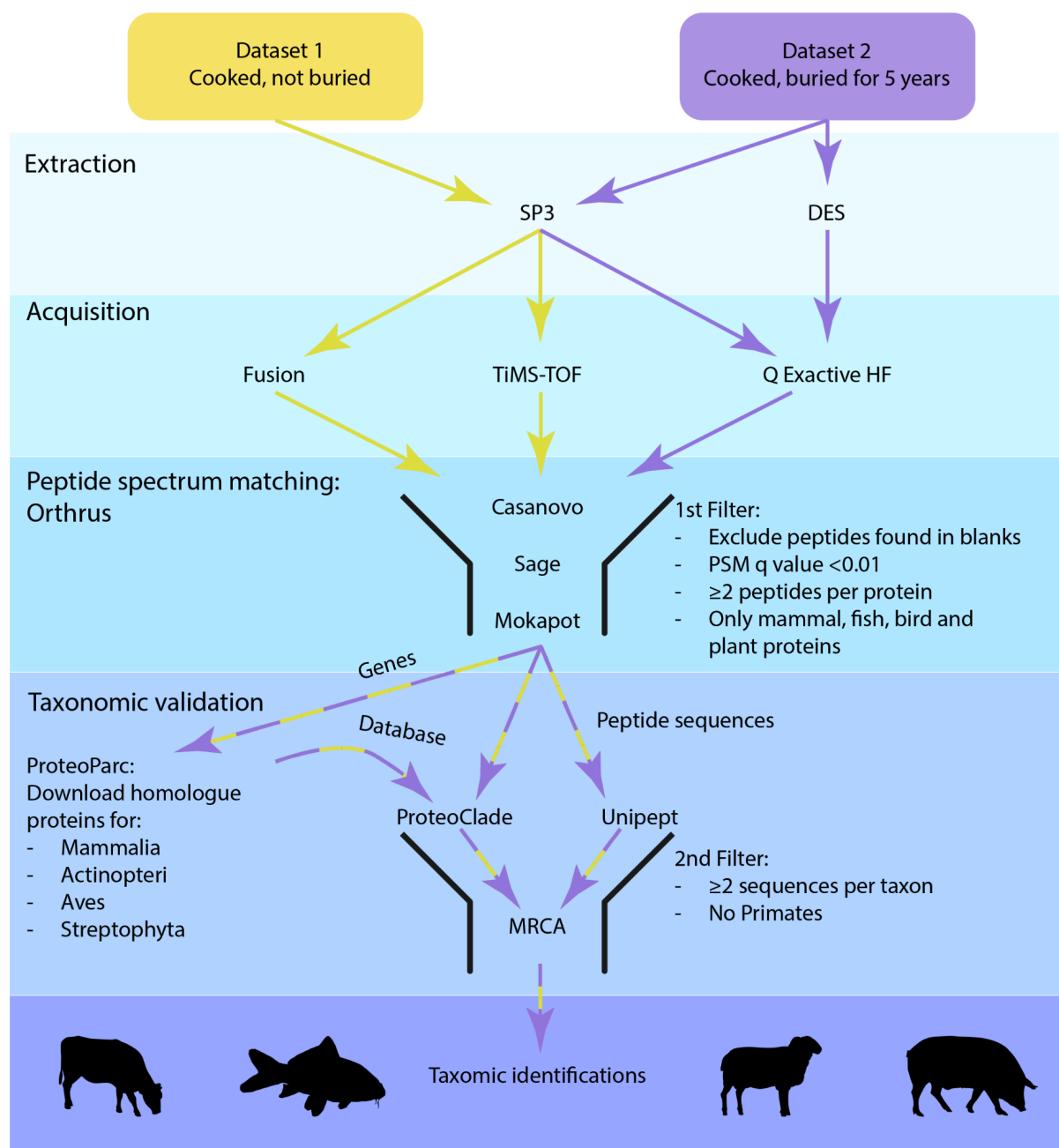


Figure 1, Overview of the different methods used in the analysis of the samples.

4.2 Materials

We analysed two sets of 10 foodcrusts each for a total of twenty samples (Table 1). The foodstuffs that were cooked to form the foodcrust, the ‘input’ of the experiment, are listed in Table 2. Of dataset 1 foodcrusts 7 and 8 were produced in 2007, while foodcrusts 1, 2, 3, 4, 5 and 6 derive from experiments performed in 2008. Sample 10, was made in 2012 and sample 9 in 2015. All foodcrusts from dataset 1 were cooked in replica ceramic vessels filled with water and heated on open fires for one to two hours. Sample 10 was buried for one year after cooking, the other samples were not.

The foodcrusts of dataset 2 were produced during experiments in November 2017 by boiling the foodstuffs in water in handmade pottery above an open fire for several hours until

the water evaporated. Following the cooking, the vessels were emptied, intentionally broken and buried at a depth of roughly 0.5 m in sandy soil. In December 2022, after 5 years of burial, the vessels were re-excavated and stored dry at room temperature. Additionally, one soil sample taken from near sample 71 was analysed. The sediment sample was labelled 71S.

Table 1, Overview of the analysed samples.

Dataset	Extraction method	Clean up	MS method	Number of samples
1	SP3	ZipTip	Fusion	10
1	SP3	EvoTip	timsTOF	10
2	SP3	EvoTip	QExactive	10
2	DES	EvoTip	QExactive	8

Table 2, List of analysed experimental foodcrusts and their original contents.

Dataset	Sample ID	Cooked foodstuffs
1	1	50% roe deer (<i>Capreolus capreolus</i>) meat, 25% rocket (<i>Diplotaxis tenuifolia/Eruca sativa</i>) and 25% chard (<i>Beta vulgaris</i>).
1	2	27% carrots (<i>Daucus carota</i> spp. <i>sativus</i>), 38% celery stalks (<i>Apium graveolens</i> var. <i>dulce</i>), 36% Brussels sprouts (<i>Brassica oleracea</i> var. <i>gemmifera</i>)
1	3	50% plaice (<i>Pleuronectes platessa</i>) flesh and 50% roe deer (<i>Capreolus capreolus</i>) meat
1	4	50% plaice (<i>Pleuronectes platessa</i>) flesh and 50% roe deer (<i>Capreolus capreolus</i>) meat
1	5	50% cod (<i>Gadus morhua</i>) flesh and 50% celery stalks (<i>Apium graveolens</i> var. <i>dulce</i>)
1	6	Porpoise (<i>Phocoena phocoena</i>) blubber
1	7	Wild boar (<i>Sus scrofa</i>) meat
1	8	Roach (<i>Rutilus rutilus</i>) flesh
1	9	Mussels (<i>Mytilidae</i>) flesh
1	10	Herring (<i>Clupea harengus</i>) flesh and spelt (<i>Triticum spelta</i>)
2	71	Cod (<i>Gadus morhua</i>) flesh
2	74	Cow (<i>Bos taurus</i>) milk
2	75	Quail (<i>Coturnix coturnix</i>) egg
2	81	Rose hips (<i>Rosa canina</i>) fruit
2	83	Bream (<i>Abramis brama</i>) meat
2	86	Wild boar (<i>Sus scrofa</i>) flesh and crab apples (<i>Malus sylvestris</i>)
2	91	Apple (<i>Malus</i> sp.) and barley (<i>Hordeum vulgare</i>)
2	97	Pike (<i>Esox lucius</i>) roe
2	98	Water chestnut (<i>Trapa natans</i>)

4.3 Methods

4.3.1 Protein extraction

Two extraction protocols were used to extract proteins. The first protocol is a modified version of the SP3 protocol (Cleland 2018) adapted to foodcrusts (Harvey and Hendy 2025). The most notable changes include a lack of demineralisation and substitution with guanidine hydrochloride incubation. The second protocol (Dekker and Ramsøe 2025) is an adaptation of the DES (Deep Eutectic Solvent) extraction protocol (Chowdhury et al. 2022).

SP3 extraction

In short, roughly 20 mg foodcrust sample, either in the form of powder or small flakes, was incubated in 150 μ L 6 M guanidine hydrochloride at 65°C for one hour. The samples were centrifuged at 12500 g for one minute and the supernatant was transferred to a new tube, where 10 μ L of a 100 mM TCEP (Tris(2-carboxyethyl)phosphine) and CAA (2-Chloroacetamide) solution was added to it. The samples were then incubated at 99°C for 10 minutes.

Subsequently the proteins were filtered using a mixture of Sera-Mag hydrophilic and hydrophobic Speedbeads (Cytiva). The filtering was done in the following way: 10 μ L of a 50 μ g/ μ L Speedbead solution was added to each sample and gently homogenised. Then 175 μ L of 100% ethanol was added to each sample and the samples were placed on a magnetic rack for around two minutes. The supernatant was then taken off and discarded and the magnetic beads were rinsed three times by adding 500, 300 and 200 μ L 80% ethanol respectively and allowing to settle on the magnetic rack for two minutes, after which the supernatant was discarded.

After the last wash the samples were incubated in 100 μ L 50 mM ammonium bicarbonate for three minutes at 37°C and 750 RPM. Afterwards, 1 μ L of a 0.2 μ g/ μ L trypsin solution was added and the samples were left to incubate for 18 hours (overnight) at 37°C and 750 RPM. After digestion samples were acidified with 10 μ L of 5% trifluoroacetic acid (TFA).

The SP3 samples that were analysed on the Fusion Tribrid Orbitrap were filtered using ZipTips. The other SP3 extracted samples were cleaned using EvoTips.

DES extraction

The DES solution was made by combining urea and guanidine hydrochloride (GuHCl) in 2:1 molar ratio and mixing this at 80°C until the mixture becomes a clear homogeneous liquid. After 20 minutes 300 μ L of the DES solution was added to 20 mg of powder or flakes of foodcrust. The samples were then immediately vortexed and incubated for four hours in a heating block at 65°C, while shaking at 750 RPM. Once the incubation was finished the samples were removed from the heating block and 0.5 mL purified water was added to them. Samples were briefly vortexed and centrifuged at 15000 rcf for 15 minutes. Then 15 μ L of a 100 mM TCEP and 100 mM CAA solution was added to each sample. Samples were homogenised and then briefly centrifuged for one minute at 15000 rcf before incubation at 99°C for 10 minutes.

Following the reduction and alkylation the samples were cleaned using C18 StageTips. In brief, the StageTips were conditioned with 150 μ L 100% methanol, followed by

an initial wash of 150 μL of 80% acetonitrile (ACN) and 0.1% TFA and a second wash of 150 μL of 0.1% TFA. Then 150 μL of sample was added to each StageTip and these were washed with two sequential washes of 150 μL of 0.1% TFA. Lastly, the proteins were eluted in 50 μL of 50% ACN + 0.1% TFA. The proteinaceous solution was then dried to nearly dry with a speedvac at 60°C and resolubilised in 100 μL 50 mM ammonium bicarbonate (ABC).

The extracted proteins were subjected to two enzymatic digestions. First, they were digested using 1 μL of 0.4 $\mu\text{g}/\mu\text{L}$ chymotrypsin (Promega) at 25°C for 18 hours. Upon completion of this first digestion, 1 μL of 0.4 $\mu\text{g}/\mu\text{L}$ trypsin (Promega) was added and the samples were incubated at 37°C for 18 hours. Following digestions the samples were acidified using 10 μL of 5% (v/v) TFA.

All DES extracted samples were cleaned up using EvoTips.

ZipTip clean up

Prior to loading the samples on the C18 ZipTips the ZipTips were primed with 100 μL of 50% ACN and 0.1% TFA and washed twice with 100 μL 0.1% TFA. 100 μL of sample was then loaded on the ZipTips and the ZipTips were washed twice again with 100 μL 0.1% TFA. Finally the peptides were eluted in 100 μL 50% ACN + 0.1% TFA.

EvoTip clean up

Prior to mass spectrometry the samples were loaded on EvoTips according to manufacturer's protocol by first loading 20 μL of 0.1% formic acid (FA) in ACN on all EvoTips. The EvoTip filters were then submerged in isopropanol for one minute and the EvoTips were washed by adding 20 μL of 0.1% FA. Then 20 μL of sample was loaded onto each EvoTip and they were washed thrice by adding 50 μL of 0.1% FA. Lastly, 100 μL of 0.1% FA was loaded on each EvoTip in order to prevent the EvoTips from drying out before acquisition.

4.3.2 Mass spectrometry data acquisition

The mass spectrometric data for Dataset 1 was acquired in the Centre for Excellence in Mass Spectrometry at the University of York. Initially this was done using an mClass nanoflow UPLC (Ultra Performance Liquid Chromatography) system (Waters) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific), but later the samples were reanalysed using an EvoSep One EV-1000 liquid chromatography system coupled with a timsTOF HT (Bruker). Dataset 2 was acquired using the EvoSep One EV-1000 coupled to a Q Exactive HF (Thermo Scientific) mass spectrometer at the Institute for Biochemistry and Molecular Biology at the University of Southern Denmark.

Orbitrap Fusion Tribrid

The analysis of Dataset 1 was performed using a mClass nanoflow UPLC (Ultra Performance Liquid Chromatography) system (Waters), which was equipped with a nanoEase M/Z Symmetry 100 Å C18, 5 μm trap column (180 μm x 20 mm, Waters) and a PepMap, 2 μm , 100 Å, C18 EasyNano nanocapillary column (75 μm x 500 mm, Thermo). The UPLC was coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) with an EasyNano ionisation source (Thermo Scientific) and the mass spectrometric data was acquired using the same settings as published in Dekker et al. ((Dekker et al. 2023). In brief, the peptides were first washed from the trap column onto the capillary column with

0.05% trifluoroacetic acid for five minutes at a flow rate of 15 $\mu\text{L}/\text{min}$. Then the flow was switched to the capillary column and for seven minutes it was washed with a mixture of solvent A (0.1% FA (formic acid)) and solvent B (0.1% FA in acetonitrile). The starting concentration of solvent B was 3% and this was steadily increased to 10% over the course of the wash. After this initial wash the concentration of solvent B was increased to 35% over 30 minutes and finally to 99% over five minutes. Lastly, 99% solvent B was used to wash the columns for a final four minutes.

For the Orbitrap Fusion Tribrid ESI-MS and MS² spectra were acquired in positive mode using Xcaliber (version 4.0, Thermo Scientific) using the following instrument settings: on spray voltage, 1,900 V; sweep gas, 0 Arb; ion transfer tube temperature; 275°C. MS¹ spectra were acquired in the Orbitrap with: resolving power: 120,000 FWHH, scan range: m/z 375-1,500; AGC target, $4e^5$; max fill time, 100 ms. Acquisition was data dependent and performed in topN mode selecting the 12 most intense precursors with charge states >1 . Easy-IC was used for internal calibration and precursor was achieved through HCD. Dynamic exclusion was performed for 50 s post precursor selection and a minimum threshold for fragmentation was set at $5e^3$. MS² spectra were acquired in the Orbitrap with: resolving power: 30,000 FWHH, max fill time, 100 ms., HCD; activation energy: 32 NCE.

timsTOF HT

The reanalysis of dataset 1 was performed using an EvoSep One nanoflow UPLC (EV-1000, EvoSep) using the 100 samples per day (SPD) gradient (from Evosep One HyStar Driver 2.4.125.0) on an EvoSep C18 Performance column (EV1109, 8 cm x 150 μm , 1.5 μm). The EvoSep was coupled to a timsTOF HT mass spectrometer (Bruker) with a CaptiveSpray ionisation source. Positive PASEF-DDA, ESI-MS and MS² spectra were acquired using Compass HyStar software (version 6.2, Bruker). Instrument source settings were: capillary voltage, 1,600 V; dry gas, 3 l/min; dry temperature; 180°C. Spectra were acquired between m/z 100-1,700. TIMS settings were: 1/K0 0.6-1.60 V.s/cm²; Ramp time, 100 ms; Ramp rate 9.42 Hz. Data dependent acquisition was performed with 10 PASEF ramps and a total cycle time of 1.17 s. An intensity threshold of 2,500 and a target intensity of 20,000 were set with active exclusion applied for 0.4 min post precursor selection. Precursor fragmentation was achieved with CID and collision energy was interpolated between 20 eV at 0.6 V.s/cm² to 59 eV at 1.6 V.s/cm².

Q Exactive HF

The samples of dataset 2 were analysed in Odense at the Institute for Biochemistry and Molecular Biology using an Evosep One liquid chromatography system (Evosep) coupled to a Q Exactive HF quadrupole-orbitrap hybrid mass spectrometer (Thermo Scientific). The Evosep system was equipped with an EV1064 column (8 cm x 100 μm , 3 μm) and was run using the 60 samples per day standard method. Mass spectrometry data was acquired with an MS¹ resolution of 120 000 and a maximum injection time of 100 ms. The Automatic gain control (AGC) was set to $3e^6$ and the scan range was 350-1600 m/z . Precursors were fragmented using HCD. For MS² a top 12 approach was used with a resolution of 15000, a maximum injection time of 100 ms and AGC set to $1e^5$. Only ions with a double, triple or quadruple charge were allowed for MS². The isolation window was 1.2 m/z and a dynamic exclusion of 10 s was maintained.

4.3.3 Bioinformatic analysis

For data analysis the MGF format files produced by the mass spectrometers were used. The files were then analysed using Orthrus (Chiang and Collins 2024), a novel protein identification tool combining the transformer-model based *de novo* identification tool Casanovo (Yilmaz et al. 2022, 2024) with Sage, a traditional protein database search engine (Lazear 2023). Sage was run tryptically allowing for three PTMs per peptide and a maximum of two missed cleavages. The selected PTMs were oxidation (M), hydroxyproline and deamidation (NQ). In order to use a protein database wide enough to include all potential taxa used in the cooking experiments, yet sufficiently restricted to prevent overinflation of the number of decoy matches, we used a database composed of all proteins identified by Casanovo during the *de novo* search for the Sage closed search. The Peptide Spectrum Matches (PSMs) identified by Orthrus were filtered to only include PSMs with a q value smaller than 0.01 and that were associated with proteins represented by at least two peptides (modified peptides were considered separate peptides).

The Orthrus results contained a wide variety of identified taxa, including bacteria, fungi and nematodes. Considering the main purpose of this study is to test the ability of proteomics to identify the cooked resources in foodcrust, we focused our downstream analysis only on proteins associated with mammal (Mammalia), bird (Aves), plants (Streptophyta) or fish (Actinopteri) taxa. We acknowledge that there might be proteins from other taxa, such as fermentation related bacteria, that might be informative for the study of diet, but which this approach excludes. However, the methodological work required for the taxonomic validation of these taxa and for the distinction between food, endogenous and soil introgressed proteins exceeds the scope of this article.

For the remaining proteins it remains necessary to validate the initially identified taxa, as protein identification tends to annotate peptides as belonging to a single species even when the sequence is shared among multiple taxa. Two approaches for taxonomic validation were tested. For the first approach we uploaded all distinct peptide sequences to the Unipept webservice (Gurdeep Singh et al. 2019) (v. 5.1.5), which matched these to the entirety of UniProt (2024.03) and provided a MRCA (Most Recent Common Ancestor) for each peptide sequence.

As for the second approach, its first step consisted of creating a custom protein database using ProteoParc (v1.1) (Martín Carrillo 2024) by using the gene names associated with all observed proteins and downloading all homologous proteins. This process was split for mammals, birds, plants and fish, i.e. that all mammalian homologues were downloaded for all observed mammalian proteins. The programme ProteoClade (Mooradian et al. 2020) was used to generate an in silico tryptic peptide database of these protein sequences, allowing for two missed cleavages. ProteoClade was then used to annotate each peptide sequence with all species that possess it. Lastly we retrieved the MRCA of each distinct peptide sequence using the ETE 3 Python package (Huerta-Cepas, Serra and Bork 2016). All taxa, excluding any Primates, represented by at least two distinct peptide sequences (of the same taxonomic level or more specific) were considered valid identifications.

4.4 Results

In general the amount of recovered dietary, i.e. proteins ascribed to a fish, bird, mammal or plant taxon, proteins is low (Fig. 2). Dataset 2 in particular yielded small numbers

of proteins and from several samples no proteins could be retrieved at all. In fact all sets of analyses had at least one sample from which no dietary proteins could be retrieved.

Upon comparing the methodological differences tested within Datasets 1 and 2, some patterns appear. Interestingly, in the results from Dataset 2 there is no overlap in which samples yielded no proteins between the DES and the SP3 analysis, nor are the samples without any detected proteins the same in the samples analysed with the Fusion and timsTOF in Dataset 1. Dataset 1 allows the comparison of the recovered proteome between the Fusion and the timsTOF. The Fusion analysis of BT3 provided the largest number of peptides among all samples of Dataset 1 and had on average the most peptides per sample, but the timsTOF data had on average the most proteins per sample.

Dataset 2 was only analysed with the QExactive HF, but instead facilitates the comparison between two extraction protocols: DES and SP3. The SP3 extraction allowed the identification of proteins from more samples than the DES extraction. However, when the DES extraction did result in identifiable proteins, it seems to yield more peptides than the SP3 extraction.

Comparing Dataset 1 and 2, Dataset 2 yielded much lower mean protein and peptide numbers, but the samples of Dataset 2 were cooked and buried, whereas only one of the samples from Dataset 1 had been buried. The burial of sample 9 from Dataset 1 was also much shorter than the burial time of the samples of Dataset 2 (1 year vs 5 years). A Kruskal-Wallis test confirms that indeed the number of identified proteins was significantly different between the different datasets ($p = 0.0007585$). Additionally, within a dataset, the difference in identified proteins was also significant between the different samples ($p = 0.00944$).

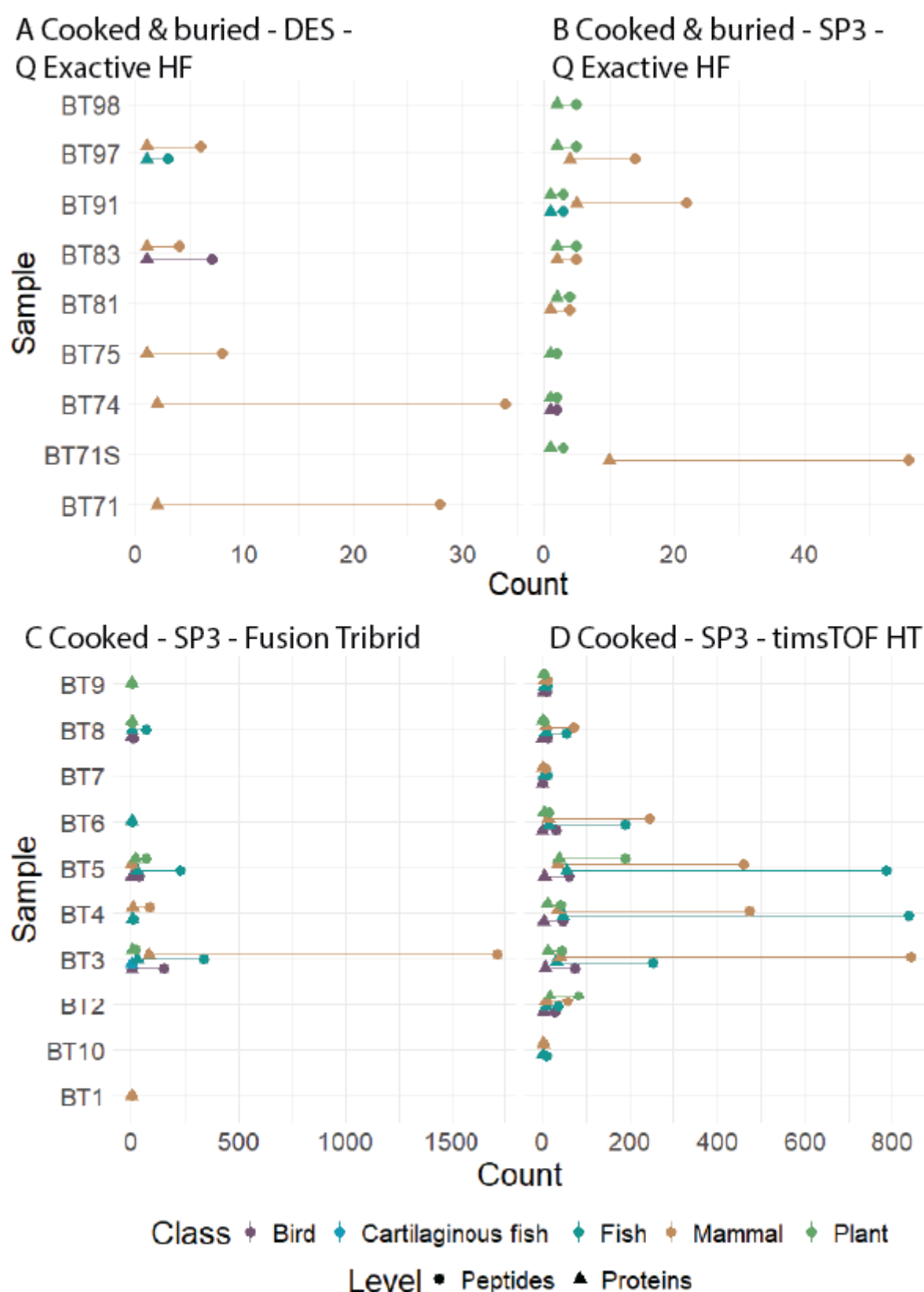


Figure 2, Number of proteins (triangles) and peptides (circles) per sample. Each sample takes up one row on the y-axis. A: Dataset 2, DES extraction, Q Exactive HF acquisition, B: Dataset 2, SP3 extraction, Q Exactive HF acquisition, C: Dataset 1, SP3 extraction, Fusion acquisition, D: Dataset 1, SP3 extraction timsTOF acquisition.

The data in Fig. 2 represents the protein identifications of the Orthrus search. The taxonomic assignment of protein search software is often initially incorrect, but this is an artefact of the way these programs operate. Part of the issue lies in the fact that the search algorithms annotate each PSM with the highest scoring database entry, which will correspond to a single species even though the peptide sequence might be shared among multiple taxa. The second part of the problem is that the employed databases are often not comprehensive and that overextending the size of the protein database reduces the amount of identified PSMs (Rodriguez Palomo et al. 2024). To obtain accurate peptide taxonomic identifications further processing is required. We made use of a custom database containing entries from 18 commonly used protein repositories (Metadata: SI3, fasta: SI4) to annotate each distinct peptide sequence with all species that share it. For each peptide sequence we then obtained the MRCA covering all matching species and tabulated the number of peptide sequences per taxon (Dataset 1: Table 4, Dataset 2: Table 5).

4.4.1 Dataset 2

The difference between the two datasets is striking. Only six taxa could be identified in the samples of Dataset 2, four of which are very unspecific. As for the specific matches, we can be quite sure that the identification of *Labeo rohita* is a mismatch, as the species is native to Southeast Asia. It is likely that the peptides matched to *Labeo rohita* instead originated from another cyprinid species. The *Sambucus nigra* identification could be genuine, as it is commonly found in Northern Europe and produces edible berries.

The number of identified sequences in Dataset 2 is also fairly low. In two out of three cases where both the SP3 and the DES analysis provided results the SP3 yielded more peptides (BT83, BT97). There was one case (BT7) where the DES analysis yielded data, but the SP3 did not and the opposite was observed in three other cases (BT71S and BT81, BT98). Samples BT86 and BT91 were only analysed using the SP3 extraction method, so for these the performance of the two extraction methods could not be compared.

For all files that yielded peptide identifications apart from BT71-SP3, ProteoClade matched more PSMs to a taxonomic group than Unipept. In general, Unipept and ProteoClade made similar taxonomic matches, except for *Labeo rohita* and *Sambucus nigra*. The two only species identifications were ProteoClade and Unipept specific.

Summarising the proteomic analysis of Dataset 2 (Table 3), it seems that samples BT71, BT71S, BT75 and BT97 only contained mammalian foodstuffs. Samples BT81 and BT86 only yielded evidence of plant food, while in the case of BT74 the DES analysis found evidence for mammal proteins and the SP3 analysis for plant proteins. From sample BT83 DES could also only recover mammal proteins, but the SP3 analysis found mammal and plant proteins. More specifically, two sequences from BT83 were identified as Fabales, which is a plant taxonomic group that includes Fabaceae (legumes), Polygalaceae, Surianaceae and Quillajaceae. Of these four families, Fabaceae seems the most likely candidate as it contains many commonly consumed plant species. Sample BT91 was the only sample which contained traces of both mammal, fish (likely cyprinid) and plant proteins.

Table 3, Initial interpretation of proteomic results versus the cooked foods (CF) for each foodcrust of Dataset 2.

Sample	DES		SP3		CF
	Proteoclade	Unipept	Proteoclade	Unipept	
BT71	Mammal meat	Mammal meat	-	-	<i>Gadus morhua</i> flesh
BT71S	-	-	Mammal unspecific	Mammal unspecific	Dirt from BT71
BT74	Laurasiatheria meat	Mammal meat	Plant	Plant	<i>Bos taurus</i> milk
BT75	Mammal meat	Mammal meat	Plant	Plant	<i>Coturnix coturnix</i> eggs
BT81	-	-	Plant	Plant	<i>Rosa canina</i> fruit
BT83	Mammal meat	Mammal meat	Mammal & legumes unspecific	<i>Sambucus nigra</i> unspecific	<i>Abramis brama</i> flesh
BT86	NA	NA	Plant greens	Plant unspecific	<i>Sus scrofa</i> meat & <i>Malus sylvestris</i> fruit
BT91	NA	NA	<i>Labeo rohita</i> skin & unspecific mammal & unspecific plant	Unspecific mammal & <i>Sambucus nigra</i>	<i>Malus</i> sp. Fruit & <i>Hordeum vulgare</i> seeds
BT97	Mammal meat	-	Unspecific mammal	Unspecific mammal & plant	<i>Esox lucius</i> roe
BT98	-	-	-	Unspecific plant	<i>Trapa natans</i> fruit

4.4.2 Dataset 1

In contrast to Dataset 2, Dataset 1 is much richer in proteins. Three samples in particular (BT3, BT4 and BT5) yielded an abundant and diverse proteome. Samples BT6 and BT8 also contained a decent number of peptide sequences, but most of these were fairly unspecific with the exception of *Gadus morhua* sequences. Some of the samples only yielded identifiable peptide sequences for one of the mass spectrometers, one sample (BT1) in the case of the Fusion and three for the timsTOF (BT2, BT7 and BT10). There are 17 species

specific matches in Table 4. Of these, *Culter alburnus*, *Cyprinus carpio*, *Danio rerio*, *Labeo rohita*, *Makaira nigricans*, *Thunnus albacares*, *Bos grunniens*, *Gossypium australe* and *Zea mays* are not native to Northern Europe and likely represent mismatches from closely related species. *Arabidopsis thaliana* is native to Northern Europe, but it is also a model organism for plant studies. Although it is safe for human consumption and should not be dismissed entirely as a possibility, there is a lack of evidence for its consumption in the archaeological record and may be a mismatch from other plant species. The remaining species level identifications are *Gadus morhua*, *Ovis aries*, *Canis lupus*, *Equus caballus*, *Rangifer tarandus*, *Sus scrofa*, *Apium graveolens* and *Oryctolagus cuniculus*. All of these, except for *Canis lupus* (BT3-4), *Gadus morhua* (BT5-6-7) and *Apium graveolens* (BT5), were solely found in sample BT3. The *Cyprinus carpio*, *Labeo rohita*, *Thunnus albacares*, *Canis lupus*, *Bos grunniens*, *Ovis aries* and *Sus scrofa* sequences were only found in ProteoClade files, while all *Gadus morhua* and *Apium graveolens* sequences were only recovered by Unipept. All other taxonomic identifications were of genus level or higher.

Table 4, Count of distinct peptide sequences per taxonomic group for all samples in Dataset 1. F denotes the data acquired with the Fusion, T data acquired with the timsTOF. PC stands for ProteoClade, UP for Unipept.

MRCA	BT1			BT2			BT3			BT4			BT5			BT6			BT7			BT8			BT9			BT10			
Sample	F	UP	T	F	UP	T	F	UP	T	F	UP	T	F	UP	T	F	UP	T	F	UP	T	F	UP	T	F	UP	T	F	UP	T	
Method	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	
Validator	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	PC	UP	
Bird							2	2	4			3	2	3			2														
Aves							2	2	4			1		1																	
Galloanserae																	2														
Neoaves												2		2																	
Neognathae												2		2																	
Fish					8	2	75	4	64	20	4	4	175	65	78	42	181	53		34	15		2	6	27	10	3	9			3
Acanthomorpha												1																			
Actinopteri																				1				1							
Clupeocephala										1											1						1				
Culter alburnus																									3						
Cyprinidae							2		1					2	2																
Cypriniformes							2					1																			
Cyprininae									1																						
Cyprinoidei					1		2		1			1	3	2			1							2							
Cyprinus carpio					3		2		9			3																			
Danio													2		2																
Danio rerio									4	3		4			2				3						4	4					
Danioninae												2																			
Fish					4	2	49	4	36	16	1	4	124	61	54	23	134	41		18	4		2		16	6	1	2			3
Gadidae															1																
Gadiformes															2		2			2				2				2			
Gadoidei															2		2			1				1				1			
Gadus															4		4														
Gadus morhua															8		8							2				3			
Labeo rohita							16		10		3	39		17		35			12						4						
Makaira nigricans												2	2																		
Percomorphaceae															2																
Siluriformes															2																
Siluroidei																	2														
Thunnus albacares							2																								
Mammal	2				12	2	638	265	185	82	26	13	82	18		17	3		59	12		3			16	7		5	3		3
Artiodactyla							14	12		2	2		1																		
Bos grunniens																															
Bos sp.							2																								
Bovidae							1		3			2																			
Bovinae							1		1	2					2																
Canis lupus												2																			
Carnivora																															
Cervidae												1	1						2												
Equus caballus							2	2																							
Laurasiatheria							27	24																							
Mammal	2				9	2	553	196	164	68	24	11	65	13		15	3		51	10		3			16	7		3	1		3
Odocoileinae							1	2					2	2																	
Oryctolagus cuniculus							3																								
Ovis aries							9																								
Pecora					2		19	28		5	5		6	2					5	2											
Rangifer tarandus							2																								
Suina							1																								
Sus scrofa							3																								
Plant						53	14	3		15	5		6	2	37	28	63	26		5	4						2				
Apiodeae															6																
Apium graveolens															2																
Arabidopsis thaliana						9																									
Brassicaceae						2																									
Fabids																															
Gossypium australe																															
Plant					44	12	3		13	5		6	2	31	20	58	24		5	4							2				
Zea mays															2		5	2													

Table 5, Count of distinct peptide sequences per taxonomic group for all samples in Dataset 2. PC stands for ProteoClade, UP for Unipept.

MRCA																					
Sample	BT71		BT71S		BT74		BT75		BT81		BT83		BT86	BT91	BT97		BT98				
Method	DES	SP3	DES	SP3	DES	SP3	DES	SP3	DES	SP3	DES	SP3	SP3	SP3	DES	SP3	DES	SP3			
Validation	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	PC UP	
Fish														3							
Labeo rohita														3							
Mammal	16	2			17	6	21	5		6	2		3	2	4		11	4	5	8	6
Laurasiatheria							2														
Mammal	16	2			17	6	21	3		6	2		3	2	4		11	4	5	8	6
Plant							6	2		4		3	3		5	5	4	2	3	4	
Fabales														2							
Plant							6	2		4		3	3		3	2	4	2	3	1	
Sambucus nigra														3					3		

The samples of Dataset 1 can be divided into three groups based on the richness of the identified proteomes. The first group is formed of samples BT1, BT7, BT9 and BT10. These four samples contain very small numbers of generally unspecific peptides. The exception are the two *Gadus morhua* and less specific Gadoidei and Gadiformes sequences in BT7. Apart from the *Gadus morhua* sequences, the most specific identification is Laurasiatheria in BT9, which is a mammalian superorder that contains among others ungulates, carnivores, bats and shrews.

The second group of samples consists of BT2, BT6 and BT8 and forms a middle group between the extremely limited protein recovery from the first group and the rich samples of the third group. These samples contain a number of seemingly species-specific fish sequences, such as *Cyprinus carpio* (BT2), *Culter alburnus* (BT8), *Danio rerio* (BT6-8), *Gadus morhua* (BT6-8) and *Labeo rohita* (BT6-8). The sequences from *Cyprinus carpio*, *Culter alburnus*, *Labeo rohita* and *Danio rerio* likely derive from another cyprinid fish with a less extensive reference proteome, as these four are not native to Northern Europe. Likewise the *Arabidopsis thaliana* sequences in BT2 probably derive from a different plant species. Although the *Cyprinus carpio*, *Labeo rohita* and *Gadus morhua* are found in the same samples, the first two species are only found in ProteoClade samples while *Gadus morhua* is only detected by Unipept. This implies that most likely only one of the two analyses is correct and that the other's misidentification is the result of a difference in the used database.

Samples BT3, BT4 and BT5 have yielded by far the most identifiable sequences of all samples and the differences between the two mass spectrometers is also greatest in these three samples. All three samples seem to contain mammal, fish, plant as well as bird peptides, although the number of plant peptides in samples BT3 and BT4 is fairly limited and was observed in all analyses. The bird peptides were recovered in even smaller numbers and their most specific identification is Neoaves in BT5. The Neoaves clade contains most modern birds, but does not include fowl, such as chickens, ducks or geese. The majority of identified bird sequences belong to muscle proteins such as creatine kinase (P11009 and P00565) and myosin (P02609). The other proteins are A0A8V1A818, A0A8V0ZVK6 and A0A8V0XER3, which are not tissue specific.

The composition of the identified fish taxa in BT3, BT4 and BT5 is very similar. The majority of the fish sequences in all three samples predominately indicates a cyprinid origin, but the presence of peptides matching to several other taxa such as *Thunnus albacares*, *Makaira nigricans*, Percomorphaceae, *Gadus morhua* and its corresponding higher taxonomic levels suggest that some of the fish identifications may be incorrect and that either one of the identified taxa is correct or that the answer lies in the evolutionary space between some of the taxa.

In contrast to fish, bird and plant peptides, the mammal sequences allow a fairly specific taxonomic identification of the used food. BT3 contains a cluster of sequences ascribed to *Bos grunniens*, *Bos* sp., Bovinae and Bovidae. The same sample also contains nine peptides belonging to *Ovis aries* (sheep), which is also part of the family Bovidae. Moreover there are a large number of Pecora peptides, which would be congruent with a *Bos taurus* or *Ovis aries* identification. The proteins matching to Pecora and more specific Pecora taxa seem to be either muscle related (e.g. myosin, Q5E946; collagen, E1BB91; hemoglobin, P213179) or unspecific. Notable by their absence are any dairy proteins. Considering dairy proteins are frequently recovered from archaeological samples (Evans et al. 2023; Fonseca, Freeman and Collins 2024), their absence in otherwise protein rich samples might actually be evidence of their absence in the food mixture. A number of other

species are also represented in BT3, such as *Equus caballus*, *Canis lupus*, *Oryctolagus*, *Rangifer tarandus* and *Sus scrofa*. Similarly, these peptides belong either to muscle proteins (hemoglobin, P21380; myosin Q076A5) or proteins without tissue specificity. Of these all but *Canis lupus* would fit the expectations of potential food sources, yet a mixture of five to six different species seems less likely. Many of the species-specific peptides were also not observed in the Unipept analysis and their numbers are fairly low. The most specific taxonomic group with a substantial amount of evidence in BT3 would be Pecora.

The most specific mammal identification in BT4 is Odocoileinae and the use of deer is congruent with the other Cervidae peptide and the six Pecora sequences. The subfamily Odocoileinae contains several American species of deer, but also *Rangifer tarandus* (reindeer), *Alces alces* (elk) and *Capreolus capreolus* (roe deer). All three are commercially available in Northern Europe today. Sample BT5 has the least mammal peptides, but no evidence of multiple taxa. The mammal component of BT5 seems to have been derived from Bovinae, but considering the number of fish and plant sequences the mammal component was likely a minor part of the mixture.

The taxonomic identifications of the proteins from each sample of Dataset 1 are summarised and contrasted to the foods that were cooked in the pots in Table 6. It is immediately apparent that in many cases it was not possible to identify the cooked foods based on the recovered proteins. Moreover, in some samples proteins were identified from foodstuffs that were not part of the original ingredient combination.

Table 6, Initial interpretation of the proteomic results of dataset 1 and the cooked foods (CF) for each experimental foodcrust.

Sample	Fusion		timsTOF		CF
	PC	UP	PC	UP	
BT1	Mammal meat	-	-	-	<i>Capreolus capreolus</i> & vegetables
BT2	-	-	Cyprinid & Pecora flesh & plant greens	Fish & mammal flesh	Vegetables
BT3	Unspecific bird, Cyprinid, bovine, <i>Sus scrofa</i> & <i>Ovis aries</i> flesh & unspecific legumes	Fish & bovine flesh	Bird, cyprinid, bovine & unspecified plant	Cyprinid & bovine flesh & unspecified plant	<i>Capreolus capreolus</i> & <i>Pleuronectes platessa</i>
BT4	Cyprinid & Laurasiatheria flesh	Fish & Laurasiatheria flesh	Neognathae, cyprinid & bovine flesh, unspecified plant	Perch-like fish, probable <i>Capreolus capreolus</i> meat and unspecified plant	<i>Capreolus capreolus</i> & <i>Pleuronectes platessa</i>
BT5	Bird & cyprinid flesh and unspecified Brassicaceae plant	<i>Gadus morhua</i> flesh & unspecified <i>Apium graveolens</i>	Neoaves, cyprinid & mammal flesh & unspecified plant	<i>Gadus morhua</i> & mammal flesh & unspecified plant	<i>Gadus morhua</i> & vegetables
BT6	-	-	Cyprinid & Pecora flesh, unspecified plant	Galloanserae, <i>Gadus morhua</i> & Pecora flesh, unspecified plant	Porpoise
BT7	-	-	Fish and mammal flesh	<i>Gadus morhua</i> flesh	<i>Sus scrofa</i>
BT8	Cyprinid flesh	Cyprinid flesh	Cyprinid and mammal flesh	<i>Gadus morhua</i> and mammal flesh	<i>Rutilus rutilus</i>
BT9	Unspecified plant	-	Laurasiatheria flesh	Laurasiatheria flesh	Mytilidae
BT10	-	-	Mammal & fish flesh	-	<i>Clupea harengus</i> & <i>Triticum spelta</i>

4.5 Discussion

4.5.1 The elephant in the room

The identification of *Canis lupus* peptides in BT3 with the timsTOF was surprising and the use of dog as an ingredient in the foodcrust seems unlikely. The process used for taxonomic validation in the article should be equivalent to the standard method of using BLAST, which is why a BLAST search was not routinely performed. Nonetheless, for the *Canis lupus* peptides we decided to search these using BLAST as well. For both of the sequences BLAST found a 100% similarity with a number of Clupeocephala fish species as well as *Canis lupus*. The reason why the fish taxa had slipped through the net of ProteoClade's taxonomic annotation lies in the way the database was constructed. We used ProteoParc to download all mammalian homologues for all mammal identified proteins. The two *Canis lupus* peptides were originally identified by Orthrus as belonging to myosin 4 (UniProt ID: Q076A5). So all mammalian homologues of the gene MYH4 were included in the database. However, all Clupeocephala proteins with the same sequence according to BLAST belong to different genes. For many of them the associated gene name is a numbered locus (e.g. LOC121607914), which explains why it was not included in the database of fish proteins even though the genes MYH6, MYHB, MYH7L, MYHZ1 and MYHA were included in the ProteoParc download of Actinopteri sequences.

The mismatch in taxonomic assignment between BLAST and ProteoClade prompted us to review all species-specific taxonomic identifications, which indeed revealed that a large majority of the species-specific peptides were shared among multiple taxa. In some cases, such as the *Labeo rohita* or *Zea mays* peptides this was expected and the revised taxonomic identifications (Table 7) do not change the interpretation of the contents of the foodcrusts. In other cases, such as the *Oryctolagus cuniculus* peptides, their reassignment to various higher taxonomic levels suggests that this species was not present in the foodcrust.

Table 7, BLAST taxonomic assignment of ProteoClade's species specific peptides.

Dataset	ProteoClade taxon	BLAST taxon	Notes	91	BT2	BT3	BT4	BT5	BT6	BT7	BT8	BT9
Fusion	Canis lupus	Euteleostomi				1						
Fusion	Culter alburnus	Actinopteri				1		1			3	
Fusion	Cyprinus carpio	Clupeocephala				2						
Fusion	Danio rerio	Danio rerio						1			4	
Fusion	Danio rerio	Euteleostomi				1						
Fusion	Equus caballus	Equus caballus				1						
Fusion	Equus caballus	Laurasiatheria	18 hits to Sus scrofa, 1 to Equus caballus			1						
Fusion	Gossypium australe	Magnoliopsida						2				
Fusion	Labeo rohita	Actinopterygii						1			1	
Fusion	Labeo rohita	Clupeocephala						3				
Fusion	Labeo rohita	Cyprinoidei						1				
Fusion	Labeo rohita	Eukaryota										1
Fusion	Labeo rohita	Euteleostomi				11	1	10			3	
Fusion	Labeo rohita	Gnathostomata				3		1				
Fusion	Labeo rohita	Otophysi				1	2					
Fusion	Labeo rohita	Teleostei						1				
Fusion	Labeo rohita	Vertebrata				1						
Fusion	Oryctolagus cuniculus	Boreoeutheria				1						
Fusion	Oryctolagus cuniculus	Euteleostomi				1						
Fusion	Oryctolagus cuniculus	Gnathostomata				1						
Fusion	Ovis aries	Amniota				1						
Fusion	Ovis aries	Artiodactyla				1						
Fusion	Ovis aries	Boreoeutheria				1						
Fusion	Ovis aries	Boreoeutheria	Pecora + 1 rodent			1						
Fusion	Ovis aries	Eukaryota				1						
Fusion	Ovis aries	Eutheria				1						
Fusion	Ovis aries	Laurasiatheria	Bovidae + Equus			1						
Fusion	Ovis aries	Pecora				1						
Fusion	Ovis aries	Theria				1						
Fusion	Rangifer tarandus	Eutheria	Matches to Cervidae and Procavia capensis			1						
Fusion	Rangifer tarandus	Rangifer tarandus				1						
Fusion	Sus scrofa	Chordata				1						
Fusion	Sus scrofa	Theria				1						
Fusion	Thunnus albacares	Euteleosteomorpha				2						
Fusion	Zea mays	Eukaryota						1				
Fusion	Zea mays	Tracheophyta						1				
SP3	Labeo rohita	Eukaryota		1								
SP3	Labeo rohita	Euteleostomi		1								
SP3	Labeo rohita	Gnathostomata		1								
TiMS-TOF	Bos grunniens	Boreoeutheria	Bos grunniens + Hippo + primate				1		1			
TiMS-TOF	Bos grunniens	Bos grunniens					1					
TiMS-TOF	Canis lupus	Euteleostomi				2	2					
TiMS-TOF	Culter alburnus	Actinopteri			1	1	1	1	1		1	
TiMS-TOF	Cyprinus carpio	Amniota			3	3	1					
TiMS-TOF	Cyprinus carpio	Clupeocephala					2					
TiMS-TOF	Cyprinus carpio	Euteleostomi				1						
TiMS-TOF	Cyprinus carpio	Eutheria				3						
TiMS-TOF	Cyprinus carpio	Gnathostomata				1						
TiMS-TOF	Cyprinus carpio	Tetrapoda				1						
TiMS-TOF	Danio rerio	Clupeocephala				1	2	1				
TiMS-TOF	Danio rerio	Danio rerio				1			1	1	1	
TiMS-TOF	Danio rerio	Euteleostomi				1			2			
TiMS-TOF	Danio rerio	Teleostei					2					
TiMS-TOF	Labeo rohita	Actinopterygii					1	1	1			
TiMS-TOF	Labeo rohita	Chordata				2						
TiMS-TOF	Labeo rohita	Clupeocephala					3	2				
TiMS-TOF	Labeo rohita	Eukaryota					6	2	1			
TiMS-TOF	Labeo rohita	Euteleostomi				8	19	14	9			
TiMS-TOF	Labeo rohita	Gnathostomata					4	5	1			
TiMS-TOF	Labeo rohita	Neopterygii					2	2				
TiMS-TOF	Labeo rohita	Osteoglossocephalai					2	3				
TiMS-TOF	Labeo rohita	Teleostei						3				
TiMS-TOF	Labeo rohita	Vertebrata					2	3				
TiMS-TOF	Rangifer tarandus	Rangifer tarandus				1	1					
TiMS-TOF	Sus scrofa	Sus scrofa					1					
TiMS-TOF	Zea mays	Mesangiospermae						4	1			
TiMS-TOF	Zea mays	Spermatophyta						1				

BLAST is often regarded as a ‘golden standard’ for taxonomic identification and its results are interpreted as more reliable than many other software tools. However, some of the taxonomic reassignments raise questions regarding the suitability of BLAST’s approach to taxonomic identifications for materials like foodcrust. Although the details differ between specific programs, the essence of software like BLAST is to compare a list of amino acid sequences with a reference database of protein sequences. Consequently, the reliability of the taxonomic identifications of such a programme primarily depends on the scale of the used database. BLAST uses NCBI’s non-redundant protein sequences database, which is an exhaustive database containing records from all branches of the tree of life. Part of its records are protein sequences translated from sequenced genomes with little to no supporting information and some of the sequences may in fact be contaminated by other organisms (Pible et al. 2014; Steinegger and Salzberg 2020). The scale of contamination in the reference sequences and especially among potentially archaeologically interesting species is difficult to assess, but a case of ruminant contamination of proteins of the genus *Jeotgalicoccus* has previously been identified (Wilkin et al. 2021). No dairy nor *Jeotgalicoccus* proteins were identified in this study, but an argument can be made that an exhaustive database spanning taxonomic kingdoms and a wide variety of proteins is not always preferable to a more restricted database containing only sequences of proteins either known or expected to be present in the sample.

Apart from the choice of software, another factor in the process of taxonomic validation that warrants further discussion is the processing of resulting taxonomic identifications. Both BLAST and ProteoClade annotate each peptide sequence with a list of *species*, because the database they use is constructed on a species-basis. In this study, the MRCA of all listed species was designated as the final taxonomic identification, but this approach may lead to an extreme, and perhaps unnecessary, loss of specificity in some cases. For example, BT3 contains one peptide (FFEHFGLSSADAIMHNDK), which BLAST identifies as a hemoglobin protein shared by *Dama dama* (fallow deer), *Capreolus capreolus* (roe deer), *Rangifer tarandus* (reindeer), *Cervus canadensis* (wapiti) and *Procavia capensis* (rock hyrax). The first four species are all part of the family Cervidae, but the inclusion of *Procavia capensis* would render the clade Eutheria, which includes all placental mammals, the MRCA. An identification of Eutheria has limited value to the archaeological interpretation, whereas a Cervidae identification would greatly strengthen the case for the use of deer meat in BT3. Considering that *Procavia capensis* is not native to Northern Europe, its exclusion seems justified, but if the sequences are homologous any other Eutherian taxon would still be a valid identification. In this particular case we would argue that the similarity between Cervidae and *Procavia capensis* in the hemoglobin amino acid sequence is due to convergent evolution rather than homology, as there are many species occupying the evolutionary space in between that do not share this peptide sequence. For the peptide to be homologous all these taxa would then have required to acquire a mutation, while it seems more likely that the mutation for this peptide only arose twice, once in Cervidae and once for *Procavia capensis*. The Cervidae versus *Procavia capensis* represents an extreme case of homology versus convergent evolution, where the evidence seems to favour convergent evolution strongly. However, the data contains more ambiguous cases as well, such as the peptide that is shared between *Bos grunniens* (yak), *Hippopotamus amphibius* (hippopotamus) and *Saimiri boliviensis* (black-capped squirrel monkey). To further complicate matters, none of these three species is likely to have been used in forming the foodcrust. The question then becomes how to reliably and systematically decide which taxa are plausible identifications and which should be excluded. Furthermore, how can that

process of validation be included in the automated processing of thousands of PSMs? This study offers no answer, but blind tests using mixtures of known composition such as this one will be vital for optimising our ability to disentangle complex mixtures such as foodcrust.

4.5.2 Chaotic mixture

Despite adjusting the taxonomic identifications based on the BLAST results the proteomically identified taxa and the actual ingredients differ significantly. As the errors can be found in both the ProteoClade, Unipept and BLAST results the source must be sought at an earlier step in the workflow. There could be a degree of contamination during the extraction process or carry over during the mass spectrometry acquisition. Extraction and machine blanks were included in this study to prevent precisely these issues, but we cannot exclude the option that some degree of cross-contamination remains. Alternatively, the error could lie in the peptide-spectrum matching process. This is more difficult to test. It is well known that different protein search engines can identify different peptides and proteins and the differences in the percentage of identified spectra can also vary significantly (Rodriguez Palomo et al. 2024). Although the specific pipeline (Orthrus) used in this study is novel, its components, Casanovo and Sage, are well trusted.

Lastly, there is the ever-present thorn of database selection. The identified peptides have been wrangled through a maximum of four databases. First, a custom database of UniProt entries based on the proteins identified by Casanovo. Secondly, a custom database of all the proteins identified by Sage was built using reference sequences from 18 public repositories (SI4). Thirdly, Unipept matched the sequences to the entirety of UniProt and finally BLAST employed the complete NCBI non-redundant protein database. Nonetheless some mismatches remain. The issue may partially be explained by the incompleteness of the databases. NCBI only holds 566 protein sequences for *Capreolus capreolus*, while *Ovis aries* has 513007 protein entries (National Library of Medicine). Consequently, regardless of the used workflow species such as *Capreolus capreolus* will likely not be identified. Similarly, *Pleuronectes platessa* could also not be identified in any of the samples even though it has 85527 associated protein records. It seems likely that all *Pleuronectes platessa* derived proteins have instead matched to Cyprinidae. The most frequently observed species-specific fish identification is the cyprinid *Labeo rohita*, which has 111824 associated protein records. Based on the magnitude of the difference in associated protein records between *Pleuronectes platessa* and *Labeo rohita* one would not expect *Pleuronectes platessa* to be completely replaced, but only 20 proteins of *Pleuronectes platessa* were included in the database used by ProteoClade, while *Labeo rohita* was represented by 102 proteins. Unfortunately, it seems that the difference in the availability of reference records is relatively larger for the observed proteins than in the total proteomes between these two species.

4.5.3 Attempted methodological improvements

The workflow employed in this study is the result of an attempt to maximise the PSM identification rate and reduce the dependency on databases. The underlying reasoning was that the prime way of increasing the specificity of taxonomic identifications, especially for conserved proteins such as collagen, would be to improve sequence coverage. Improving sequence coverage requires identifying additional peptides, which at the root requires a higher percentage of identified spectra. Software such as Orthrus is precisely designed to achieve this. Furthermore, by incorporating *de novo* tools like Casanovo, a measure of

independence from databases was sought, yet in order to obtain confidence in the *de novo*-identified sequences they still need to be matched against a database. It may be that, in cases where the 'true' identity of the extracted proteins is not in the database, our endeavour to assign a peptide sequence to a larger share of the observed peaks could result in increased pressure to shoehorn the peaks into a match with a similar, but different peptide sequence. The fact that the analysed samples are foodcrusts likely exacerbates the issue, as cooking-induced damage and modifications complicate identification (Bax et al. 2012; Barker et al. 2018; Hu et al. 2017; Yu et al. 2017).

Apart from maximising PSM identifications we applied several tools (ProteoParc, ProteoClade, Unipept) to validate the taxonomic identifications. As previously discussed we largely did not succeed in that endeavour. Nonetheless, we believe that these tools remain valuable additions and would advocate for their wider application. The cause of their underperformance lies not in the tools themselves, but in the state of annotation of the public repositories of protein sequences. There is little an individual researcher can do about this matter, but initiatives such as the Earth Biogenome Project (Earth Biogenome Project 2022) and Darwin Tree of Life (Darwin Tree of Life) will slowly but surely increase the coverage of reference proteomes across all taxa. Additionally, studies targeting specific studies for biomarkers (Dierickx et al. 2022; Speller et al. 2016; Peters et al. 2021) or establishing new reference sequences (Welker et al. 2015; Dekker et al. 2024; Le Meillour et al. 2020) are invaluable for advancement of palaeoproteomics as a field. As the references on the public repositories will improve, so will the power of validation tools like ProteoClade and Unipept.

Apart from validating the taxonomic specificity of the identified peptide sequences, increased scrutiny of the identified PSMs might improve the taxonomic identifications as well. In this study Mokapot, a machine-learning Python package (Fondrie and Noble 2021), was used to rescore the PSMs and all PSMs with a q-value above 0.01 were rejected. However, more stringent criteria in accepting PSMs could help to remove false positives at the start of the process facilitating the taxonomic validation of the remaining peptide sequences. PSM quality could be evaluated by setting a minimum number for the number of fragment ions needed to support a peptide sequence. A score threshold could also be employed, but the scoring system of protein database searches is complex and opaque. Any threshold might not be transferable between datasets or studies and could be arbitrary. Alternatively, modelling the expected intensity of the fragment ions of a precursor could also increase the confidence in peptide identifications (Degroeve and Martens 2013).

4.5.4 Returning to order

Precise taxonomic identifications are one of the prime aims of the palaeoproteomic investigation of organic residues, but of greater importance is that the presented taxonomic identifications are accurate. Ideally, a combination of both is achieved, but the previously presented results are neither. To align the ingredient input with the observed proteins we grouped the taxonomic identifications of dataset 2 by order level and raised the threshold of acceptance from two sequences per taxon to six (Table 8). Dataset 1 was not reordered, as the original identifications presented in Table 5 are already on order level or higher, except for the *Labeo rohita* and *Sambucus nigra* identifications.

Table 8, Sequence taxonomic identifications of dataset 1 grouped at order level with a minimum six sequences.

MRCA Sample Method Validation	BT1			BT2			BT3			BT4			BT5			BT6			BT7			BT8			BT9			BT10		
	F		T	F		T	F		T	F		T	F		T	F		T	F		T	F		T	F		T	F		T
	PC	UP		PC	UP		PC	UP		PC	UP		PC	UP		PC	UP		PC	UP		PC	UP		PC	UP		PC	UP	
Fish							73	62	16		172	61	76	38	177	51		34	8				27	6						
Cypriniformes							24	26			48		22	43				16				11								
Fish							49	36	16		124	61	54	23	134	41		18				16	6							
Gadiformes													15	10				8												
Mammal				9	632	263	183	77	24	11	77	13		15			51	10					16	7						
Artiodactyla					52	43	13	9			12																			
Laurasiatheria					27	24	6																							
Mammal				9	553	196	164	68	24	11	65	13		15			51	10					16	7						
Plant				53	12		13			6	31	28	58	24																
Apiales												8																		
Brassicales				9																										
Plant				44	12		13			6	31	20	58	24																

The refiltered data conflicts less with the food input, although in BT2 (timsTOF), BT3 (timsTOF), BT4 (timsTOF), BT5 (timsTOF), BT6 (timsTOF) and BT8 (timsTOF) there are still taxonomic identifications from a food category (i.e. mammal, fish, plant) that do not match the input. Moreover, even though the taxa are grouped by order, *Pleuronectes platessa* is still misidentified as Cypriniformes, but this is likely a database issue. It is remarkable that all the identifications from the Fusion do match the food category, although in the case of BT6 this might be somewhat misleading as no taxonomic group of the Fusion exceeded the sequence threshold. Additionally, in the case of BT2, BT3, BT4 and BT5 only the ProteoClade output provides misidentifications, highlighting that the issue for these files might once again be in the used database.

The remaining identifications are in agreement with the known input, but they are limited in their specificity. For the samples containing *Capreolus capreolus* the identification is limited to Artiodactyla, even-toed ungulates, at best in BT3 and BT4. A positive consequence of the Artiodactyla grouping is that incorrect identifications such as Bovidae and *Ovis aries* are now merged in a less precise, but correct taxonomic placement. However, the other side of this coin is demonstrated by the Gadiformes order. *Gadus morhua* was the only ingredient that was correctly identified to species level in BT5 and BT6. Reducing the identification to order level results in at least 600 other Gadiformes species as possible sources. Upon filtering this list to a specific ecosystem, such as the Baltic Sea, there remain 13 species apart from *Gadus morhua* as valid identifications (Froese and Pauly 06-2024). In short, strict identification criteria and order-level grouping allows for acceptable accuracy in taxonomic identifications, given database-induced limitations, but the results presented here clearly suggest that a combination of precise and accurate identifications remains out of reach at the moment.

4.5.5 Implications for archaeological research

The limited precision of the taxonomic identifications reduces the value of palaeoproteomic analysis in comparison to other more robust methods for investigating mixtures like foodcrust. In lipid analysis a distinction can be made between ruminant adipose fats and porcine fats, but if grouped by order both of these taxa would be assigned to Artiodactyla, as would be most other commonly exploited herbivore species. The taxonomic resolution of proteomic analysis would then effectively only allow distinguishing between even- and odd-toed ungulates, carnivores and elephants.

In the case of fish and plants the taxonomic resolution would likely be lower. Even when grouped by order the *Pleuronectes platessa* proteins were not matching to the correct taxon. The *Gadus morhua* proteins were correctly identified by Unipept, but it is not sure how easy to identify proteins from other fish taxa would be. As for plants, most plant proteins

were limited to the level of 'plant', which groups all taxa between Viridiplantae and Streptophyta. It was one of the hopes of this study that a greater taxonomic resolution for plants could be achieved, as plant resources often receive less attention than the faunal component of the diet, but this proved challenging. Nonetheless, two out of three samples of dataset 2 containing plant foods yielded plant proteins, highlighting that proteomics has a fair chance of detecting plant resources if present.

The standard criteria for the identification of proteins in archaeological samples are that at least two peptides should be present per protein (Hendy et al. 2018a). However, using those criteria a large number of incorrect taxonomic identifications would have to be accepted. Therefore, at least for complex mixtures, stricter criteria seem in order. Increasing the threshold to six distinct sequences per taxon removed many of the incorrect taxonomic identifications, but requiring larger numbers of peptides may complicate the application to archaeological samples. Modern experimentally generated foodcrusts already yield limited numbers of proteins and peptides, but archaeological samples are expected to be more depleted in protein content. Requiring at least six sequences per taxon may therefore result in a significant share of samples not yielding any taxonomic identifications.

4.6 Conclusion

The aim of this study was to analyse the proteins from a set of foodcrusts with known composition and to establish the limitations of our ability to correctly identify the resources used to produce the foodcrust. As we were intending to inform our interpretations of archaeological material using this study's results we emphasised the accuracy of taxonomic identifications and went beyond standard palaeoproteomic identification criteria and methodology to establish the foodcrusts' contents. The results have been humbling. The errors in taxonomic assignment were significant and would not have been discovered had the composition of these foodcrusts been unknown. By tightening the criteria for accepting taxonomic identifications we have been able to remove most of the incorrect taxa, but this has come at the loss of taxonomic resolution. There is a place for proteomics in organic residue analysis, yet this study makes the need for methodological innovation and caution in interpretation abundantly clear.

References

- Admiraal, M. et al. (2020). Leftovers: The presence of manufacture-derived aquatic lipids in Alaskan pottery. *Archaeometry*, 62 (2), Wiley., pp.346–361. [Online]. Available at: doi:10.1111/arcm.12515.
- Barker, A. et al. (2012). An Optimized Approach for Protein Residue Extraction and Identification from Ceramics After Cooking. *Journal of Archaeological Method and Theory*, 19 (3), pp.407–439. [Online]. Available at: doi:10.1007/s10816-011-9120-5.
- Barker, A. et al. (2018). Taphonomy and negative results: An integrated approach to ceramic-bound protein residue analysis. *Journal of archaeological science*, 94, pp.32–43. [Online]. Available at: doi:10.1016/j.jas.2018.03.004.
- Bax, M.-L. et al. (2012). Cooking temperature is a key determinant of in vitro meat protein digestion rate: investigation of underlying mechanisms. *Journal of agricultural and food chemistry*, 60 (10), pp.2569–2576. [Online]. Available at: doi:10.1021/jf205280y.
- Bondetti, M. et al. (2021). Investigating the formation and diagnostic value of ω -(o-alkylphenyl)alkanoic acids in ancient pottery. *Archaeometry*, 63 (3), pp.594–608. [Online]. Available at: doi:10.1111/arcm.12631.
- Chiang, Y. and Collins, M. J. (2024). Orthrus: an AI-powered, cloud-ready, and open-source hybrid approach for metaproteomics. *bioRxiv*. [Online]. Available at: doi:10.1101/2024.11.15.623814.
- Chowdhury, M. et al. (2022). Novel Deep Eutectic Solvent-Based Protein Extraction Method for Pottery Residues and Archeological Implications. *Journal of proteome research*, 21 (11), American Chemical Society., pp.2619–2634. [Online]. Available at: doi:10.1021/acs.jproteome.2c00340.
- Chowdhury, M., Campbell, S. and Buckley, M. (2021). Proteomic analysis of archaeological ceramics from Tell Khaiber, southern Iraq. *Journal of archaeological science*, 132, p.105414. [Online]. Available at: doi:10.1016/j.jas.2021.105414.
- Cleland, T. P. (2018). Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. *Journal of proteome research*, 17 (11), pp.3976–3983. [Online]. Available at: doi:10.1021/acs.jproteome.8b00637.
- Colonese, A. C. et al. (2017). New criteria for the molecular identification of cereal grains associated with archaeological artefacts. *Scientific reports*, 7 (1), p.6633. [Online]. Available at: doi:10.1038/s41598-017-06390-x.
- Copley, M. S. et al. (2004). Organic residue evidence for the processing of marine animal products in pottery vessels from the pre-colonial archaeological site of Kasteelberg D east, South Africa. *South African Journal of Science*, 100 (5), journals.co.za., pp.279–283. [Online]. Available at: doi:10.10520/ejc96255.
- Craig, O. E. et al. (2011). Ancient lipids reveal continuity in culinary practices across the transition to agriculture in Northern Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (44), pp.17910–17915. [Online]. Available at: doi:10.1073/pnas.1107202108.
- Craig, O. E. et al. (2013). Earliest evidence for the use of pottery. *Nature*, Nature Publishing Group. [Online]. Available at: doi:10.1038/nature12109.

Craig, O. E. and Collins, M. J. (2002). The Removal of Protein from Mineral Surfaces: Implications for Residue Analysis of Archaeological Materials. *Journal of archaeological science*, 29 (10), pp.1077–1082. [Online]. Available at: doi:10.1006/jasc.2001.0757.

Cubas, M. et al. (2023). Organic Residue Analysis of Archaeological Pottery. In: *Reference Module in Social Sciences*. Elsevier. [Online]. Available at: doi:10.1016/B978-0-323-90799-6.00047-1.

Dallongeville, S. et al. (2011). Dealing with the identification of protein species in ancient amphorae. *Analytical and bioanalytical chemistry*, 399 (9), pp.3053–3063. [Online]. Available at: doi:10.1007/s00216-010-4218-2.

Darwin Tree of Life. *Darwin Tree of Life*. [Online]. Available at: <https://www.darwintreeoflife.org/> [Accessed 19 September 2024].

Degroeve, S. and Martens, L. (2013). MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics*, 29 (24), academic.oup.com., pp.3199–3203. [Online]. Available at: doi:10.1093/bioinformatics/btt544.

Dekker, J. et al. (2023). Spatial analysis of the ancient proteome of archeological teeth using mass spectrometry imaging. *Rapid communications in mass spectrometry: RCM*, 37 (8), p.e9486. [Online]. Available at: doi:10.1002/rcm.9486.

Dekker, J. A. A. et al. (2024). Palaeoproteomic identification of a whale bone tool from Bronze Age Heiloo, the Netherlands. *Peer Community Journal*, 4, peercommunityjournal.org. [Online]. Available at: doi:10.24072/pcjournal.451.

Dekker, J. and Ramsøe, M. (2025). *DES foodcrust protein extraction protocol*. [Online]. Available at: doi:10.17504/protocols.io.j8nlk9r6wv5r/v1 [Accessed 17 February 2025].

Dierickx, K. et al. (2022). Peptide mass fingerprinting of preserved collagen in archaeological fish bones for the identification of flatfish in European waters. *Royal Society open science*, 9 (7), royalsocietypublishing.org., p.220149. [Online]. Available at: doi:10.1098/rsos.220149.

Drieu, L. et al. (2020). Is it possible to identify ancient wine production using biomolecular approaches? *STAR: Science & Technology of Archaeological Research*, 6 (1), Routledge., pp.16–29. [Online]. Available at: doi:10.1080/20548923.2020.1738728.

Earth Biogenome Project. (2022). *Earth Biogenome Project*. [Online]. Available at: <https://www.earthbiogenome.org/> [Accessed 19 September 2024].

Evans, M. et al. (2023). Detection of dairy products from multiple taxa in Late Neolithic pottery from Poland: an integrated biomolecular approach. *Royal Society open science*, 10 (3), p.230124. [Online]. Available at: doi:10.1098/rsos.230124.

Evans, M. et al. (2024). The impact of cooking and burial on proteins: a characterisation of experimental foodcrusts and ceramics. *Royal Society open science*, 11 (240610). [Online]. Available at: doi:10.1098/rsos.240610.

Evershed, R. P. et al. (2008). Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature*, 455 (7212), pp.528–531. [Online]. Available at: doi:10.1038/nature07180.

Fondrie, W. E. and Noble, W. S. (2021). Mokapot: Fast and flexible semisupervised learning for peptide detection. *Journal of proteome research*, 20 (4), American Chemical Society (ACS), pp.1966–1971. [Online]. Available at: doi:10.1021/acs.jproteome.0c01010 [Accessed

24 July 2024].

Fonseca, B., Freeman, C. L. and Collins, M. J. (2024). The survival of β -lactoglobulin peptides in the archaeological record: Vulnerability vs. Sequence variation. *bioRxiv*, p.2024.09.13.612646. [Online]. Available at: doi:10.1101/2024.09.13.612646 [Accessed 17 September 2024].

Forte, V., Nunziante Cesaro, S. and Medeghini, L. (2018). Cooking traces on Copper Age pottery from central Italy: An integrated approach comprising use wear analysis, spectroscopic analysis and experimental archaeology. *Journal of archaeological science, reports*, 18, Elsevier BV., pp.121–138. [Online]. Available at: doi:10.1016/j.jasrep.2017.12.052.

Froese, R. and Pauly, D. (06-2024). *FishBase*. [Online]. Available at: www.fishbase.org [Accessed 19 September 2024].

Gurdeep Singh, R. et al. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of proteome research*, 18 (2), pp.606–615. [Online]. Available at: doi:10.1021/acs.jproteome.8b00716.

Harvey, V. L. and Hendy, J. (2025). *SP3 protocol optimised for foodcrust protein extraction from archaeological cooking vessels*. [Online]. Available at: <https://www.protocols.io/view/sp3-protocol-optimised-for-foodcrust-protein-extra-dyge7tte.pdf> [Accessed 31 January 2025].

Hendy, J. et al. (2018a). A guide to ancient protein studies. *Nature ecology & evolution*, 2 (5), nature.com., pp.791–799. [Online]. Available at: doi:10.1038/s41559-018-0510-x.

Hendy, J. et al. (2018b). Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nature communications*, 9 (1), p.4064. [Online]. Available at: doi:10.1038/s41467-018-06335-6.

Huerta-Cepas, J., Serra, F. and Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Molecular biology and evolution*, 33 (6), Oxford University Press (OUP)., pp.1635–1638. [Online]. Available at: doi:10.1093/molbev/msw046 [Accessed 6 September 2024].

Hu, L. et al. (2017). Proteomic study of the effect of different cooking methods on protein oxidation in fish fillets. *RSC advances*, 7 (44), The Royal Society of Chemistry., pp.27496–27505. [Online]. Available at: doi:10.1039/C7RA03408C [Accessed 12 November 2021].

Lazear, M. R. (2023). Sage: An open-source tool for fast proteomics searching and quantification at scale. *Journal of proteome research*, 22 (11), American Chemical Society (ACS)., pp.3652–3659. [Online]. Available at: doi:10.1021/acs.jproteome.3c00486 [Accessed 24 July 2024].

Le Meillour, L. et al. (2020). Palaeoproteomics gives new insight into early southern African pastoralism. *Scientific reports*, 10 (1), Nature Publishing Group., pp.1–11. [Online]. Available at: doi:10.1038/s41598-020-71374-3.

Lucquin, A. et al. (2016). Ancient lipids document continuity in the use of early hunter–gatherer pottery through 9,000 years of Japanese prehistory. *Proceedings of the National Academy of Sciences*, 113 (15), pp.3991–3996. [Online]. Available at: doi:10.1073/pnas.1522908113.

- Lundy, J. et al. (2024). Culinary continuity in central Japan across the transition to agriculture. *Archaeological and anthropological sciences*, 16 (7), Springer Science and Business Media LLC., p.97. [Online]. Available at: doi:10.1007/s12520-024-01992-9.
- Lyu, N. et al. (2024). Characterization of Pottery Foodcrusts Through Lipid and Proteomic Analyses: A Case Study from the Xiawan Site in Yixing City, East China. *Journal of archaeological science*, 161 (105902), papers.ssrn.com. [Online]. Available at: doi:10.1016/j.jas.2023.105902 [Accessed 8 September 2023].
- Martín Carrillo, G. (2024). *proteoparc: A bioinformatic tool to generate reference databases for ancient protein data analysis*. Github. [Online]. Available at: <https://github.com/guillecarrillo/proteoparc> [Accessed 27 October 2024].
- Montalvo-Cabrera, J. A. et al. (2024). Unraveling island economies through organic residue analysis: The case of Mocha Island (southern Chile). *Latin American antiquity*, Cambridge University Press., pp.1–15. [Online]. Available at: doi:10.1017/laq.2023.58.
- Mooradian, A. D. et al. (2020). ProteoClade: A taxonomic toolkit for multi-species and metaproteomic analysis. *PLoS computational biology*, 16 (3), p.e1007741. [Online]. Available at: doi:10.1371/journal.pcbi.1007741.
- National Library of Medicine. *NCBI*. [Online]. Available at: <https://www.ncbi.nlm.nih.gov/> [Accessed 18 September 2024].
- Peters, C. et al. (2021). Species identification of Australian marsupials using collagen fingerprinting. *Royal Society open science*, 8 (10), royalsocietypublishing.org., p.211229. [Online]. Available at: doi:10.1098/rsos.211229.
- Pible, O. et al. (2014). The importance of recognizing and reporting sequence database contamination for proteomics. *EuPA Open Proteomics*, 3, Elsevier., pp.246–249. [Online]. Available at: doi:10.1016/j.euprot.2014.04.001.
- Rodriguez Palomo, I. et al. (2024). Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *Peer Community J.*, (e107), p.2023.12.15.571577. [Online]. Available at: doi:10.24072/pcjournal.491 [Accessed 29 November 2024].
- Salque, M. et al. (2013). Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature*, 493 (7433), Springer Science and Business Media LLC., pp.522–525. [Online]. Available at: doi:10.1038/nature11698 [Accessed 18 July 2024].
- Shevchenko, A. et al. (2018). Fine Endmesolithic fish caviar meal discovered by proteomics in foodcrusts from archaeological site Friesack 4 (Brandenburg, Germany). *PloS one*, 13 (11), p.e0206483. [Online]. Available at: doi:10.1371/journal.pone.0206483.
- Solazzo, C. et al. (2008). Identification of protein remains in archaeological potsherds by proteomics. *Analytical chemistry*, 80 (12), pp.4590–4597. [Online]. Available at: doi:10.1021/ac800515v.
- Speller, C. et al. (2016). Barcoding the largest animals on Earth: ongoing challenges and molecular solutions in the taxonomic identification of ancient cetaceans. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371 (1702). [Online]. Available at: doi:10.1098/rstb.2015.0332.
- Steinegger, M. and Salzberg, S. L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome biology*, 21 (1),

p.115. [Online]. Available at: doi:10.1186/s13059-020-02023-1.

Tanasi, D. et al. (2021). Paleoproteomic profiling of organic residues on prehistoric pottery from Malta. *Amino acids*, 53 (2), pp.295–312. [Online]. Available at: doi:10.1007/s00726-021-02946-4.

Van Gijn, A. et al. (2020). Studying the life history of vessels: Creating a reference collection for microwear studies of pottery. *Pots and Practices*, research.rug.nl. [Online]. Available at: https://research.rug.nl/files/157951761/VanGijn_et_al_2020_Life_history.pdf.

Welker, F. et al. (2015). Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature*, 522 (7554), nature.com., pp.81–84. [Online]. Available at: doi:10.1038/nature14249.

Wilkin, S. et al. (2021). Dairying enabled Early Bronze Age Yamnaya steppe expansions. *Nature*, 598 (7882), Springer Science and Business Media LLC., pp.629–633. [Online]. Available at: doi:10.1038/s41586-021-03798-4 [Accessed 9 September 2024].

Yilmaz, M. et al. (2022). De novo mass spectrometry peptide sequencing with a transformer model. In: *International Conference on Machine Learning*. 28 June 2022. PMLR. pp.25514–25522. [Online]. Available at: <https://proceedings.mlr.press/v162/yilmaz22a.html> [Accessed 6 September 2024].

Yilmaz, M. et al. (2024). Sequence-to-sequence translation from mass spectra to peptides with a transformer model. *Nature communications*, 15 (1), Springer Science and Business Media LLC., p.6427. [Online]. Available at: doi:10.1038/s41467-024-49731-x [Accessed 6 September 2024].

Yu, T.-Y. et al. (2017). Cooking-induced protein modifications in meat. *Comprehensive Reviews in Food Science and Food Safety*, 16 (1), Wiley., pp.141–159. [Online]. Available at: doi:10.1111/1541-4337.12243.

5 Palaeoproteomic, lipid and SEM analysis of foodcrusts from Syltholm II (MLF00906-I/II), Denmark, reveals dairy consumption in the Mesolithic Ertebølle

Joannes Adrianus Antonius Dekker^{1,2}, Lara Gonzalez Carretero¹, Alexandre Lucquin¹, Ann-Katrin Meyer³, Yun Chiang^{2,4}, Max Ramsøe², Jasmine Lundy¹, Jonica Ella Doliente^{1,5,6}, Harry K. Robson¹⁰, Guillermo Rangel Piñeros², Chloë Baldreki⁷, Adam Dowle⁷, Tina Ravnsborg⁸, Ole Jensen⁸, Bente Philippsen⁹, Daniel Groß¹⁰, Oliver Craig¹, Matthew Collins^{2,11}, Jessica Hendy¹.

1. BioArch, University of York, UK.
2. Section for Geobiology, Globe Institute, University of Copenhagen, Denmark.
3. Archäologisches Institut, University of Hamburg, Germany?
4. The Nice Institute of Chemistry, Université Côte d'Azur, France.
5. Department of Chemistry, University of York, UK.
6. Department of Prehistory, Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona, Spain.
7. Department of Biology, Bioscience Technology Facility, University of York, UK.
8. Institut for Biokemi og Molekylær Biologi, University of Southern Denmark, Denmark.
9. National Laboratory for Age Determination, Norwegian University of Science and Technology, Norway.
10. Museum Lolland-Falster, Denmark.
11. McDonald Institute for Archaeological Research, University of Cambridge, UK.

Written with intent of submission in the journal Science Advances. Intended date of submission is 30-03-2025.

CRedit statement

J.A.A.D: Conceptualisation, Methodology, Validation, Formal analysis, Investigation, Investigation, Writing - Original draft, Writing - Review & editing, Visualisation, Project administration. **L.G.C:** Methodology, Formal analysis, Investigation, Resources, Writing - Original draft, Visualisation. **A.L:** Methodology, Formal analysis, Investigation, Resources, Writing - Original draft, Visualisation. **A-K.M:** Formal analysis, Investigation. **Y.C:** Software, Validation. **M.R:** Methodology, Investigation, Resources. **J.L:** Investigation. **J.E.D:** Methodology. **H.R:** Resources, Writing - Review & editing. **G.R.P:** Software. **C.B:** Methodology, Validation, Investigation, Resources, Writing - Original draft. **A.D:** Resources, Supervision, Writing - Review & editing. **T.R:** Validation, Investigation, Resources, Writing - Original draft. **O.J:** Resources, Supervision, Project administration. **B.P:** Resources, Data curation. **D.G:** Conceptualisation, Resources, Data curation, Writing- Review & editing, Funding acquisition. **O.C:** Resources, Supervision, Funding acquisition, Writing - Review & editing. **M.C:** Conceptualisation, Resources, Writing - Review & editing, Supervision, Project administration, Funding acquisition. **J.H:** Conceptualisation, Resources, Writing - Review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgements and funding sources

We would like to thank the Bioscience Technology Facility and Chemistry Department at the University of York for mass spectrometry access and support. The instrumentation is part of the York Centre of Excellence in Mass Spectrometry. The centre was created thanks to a major capital investment through Science City York, supported by Yorkshire Forward with funds from the Northern Way Initiative and subsequent support from EPSRC (EP/K039660/1; EP/M028127/1).

J.D. is funded by a Marie Skłodowska-Curie grant (grant agreement No. 956351) from the EU Horizon 2020 programme.

Abstract

Foodcrusts are a promising material for dietary analysis as they are direct evidence of the heat processing of foodstuffs and have the potential to provide information on both the plant and faunal component of the diet. Common methods to analyse foodcrust include lipid analysis, scanning electron microscopy (SEM) and recently proteomics, but lipid analysis is biased towards lipid-rich foods, while SEM is mainly used for plant resources and proteomics favours protein-rich foods. By combining all three methods we highlight how the methods complement and contrast each other and can provide a more comprehensive understanding of dietary practices in the past. We apply this approach to Mesolithic (Ertebølle) and Neolithic (Funnelbeaker) foodcrusts from Syltholm II (MLF00906-I/II), Denmark, to study dietary change or continuity during the Mesolithic-Neolithic transition. Our results indicate that marine resources, ruminant fats, wild tubers and dairy were exploited both during the Ertebølle and in the Funnelbeaker. Additionally, a number of Funnelbeaker foodcrusts contains evidence of cereal consumption. The evidence for dairy consumption in the Ertebølle stems from both lipid analysis as well as the presence of casein proteins. The only clear change in diet between the Mesolithic and Neolithic observed in this study is the introduction of cereals, which were only observed in the SEM analysis. The fact that cereals were only observed with the SEM highlights a broader pattern of some overlap in identified resources between lipid and protein analysis, but foodstuffs identified with the SEM were largely not recovered using the other methods.

Significance statement

The Mesolithic-Neolithic transition in Western Europe is associated with large changes in material cultures, genetic ancestry and diet. As domesticates are introduced with Neolithic farmers a shift away from wild resources, including wild plants and marine resources, is often observed. However, organic residue analysis on Neolithic cultures from the circum-Baltic indicate a different narrative, with marine resources continuing to be exploited intensively during the Early Neolithic. We now add to the growing body of evidence for dairy consumption during the Mesolithic as well, indicating a greater degree of continuity in diet throughout the transition and resource exchange between neighbouring Mesolithic and Neolithic communities.

5.1 Introduction

The change from hunter-gatherer societies to food producing economies has had an immense impact on human societies and the nature and speed of this transition remain the focus of ongoing research. In Western Europe this change is equated to the Mesolithic-Neolithic transition and comes paired with the introduction of ceramics, domesticates and a significant change in genetic makeup of the population. Denmark plays an exceptional role in the narrative of the Mesolithic-Neolithic transition in Europe. Its first Neolithic culture, Funnelbeaker (TRB), arrives relatively late (ca. 4000 BC), despite being already present in neighbouring regions centuries before (Sørensen and Karg 2014; Jensen and Sørensen 2023). Additionally, the final Mesolithic culture in Denmark, the Ertebølle (EBK), adopted pottery while retaining a hunter-gatherer subsistence strategy, similarly to other ceramic hunter-gatherer cultures on the Baltic coast (Courel et al. 2020). There is evidence of some exchange of goods between the EBK and the TRB, as well as between other Mesolithic and Neolithic cultures (Marchand 2007), such as in the form of imported axes (Klassen 2004; Hartz, Lübke and Terberger 2007), but there has been no convincing evidence for the exchange of domesticates. There have been reports of cereals (Sørensen and Karg 2014) and pig in the EBK (Krause-Kyora et al. 2013), but these are not generally accepted (Rowley-Conwy and Zeder 2014). The strongest case for the use of domesticates in the EBK comes from the detection of dairy-congruent ^{13}C isotope ratios detected during the organic residue analysis of EBK vessels (Lucquin et al. 2023). In this study we complement the case for dairy consumption in the EBK with the detection of *Bos taurus* dairy proteins in foodcrusts on late EBK vessels from Syltholm II (MLF00906-I/II), Lolland, Denmark.

The site belongs to a cluster of site complexes found as part of the Femern project. Syltholm II is part of site complex 5, which contains sites dating from the Late Palaeolithic to the Early Bronze Age. Syltholm II itself covers the Late Mesolithic to Neolithic and contains faunal remains, lithics as well as EBK and TRB pottery finds (Måge, Groß and Kanstrup 2023). As there is no clear stratigraphical separation between a Mesolithic or Neolithic layer at the site, the assignment of the foodcrust samples to the EBK or TRB is solely based on typological and technological analysis as detailed in section 5.4.1, which divided the assemblage into clear EBK, TRB and probable EBK and TRB sherds. In short, U-technique coiling, thick wall size (>10 mm) and medium temper grain size characterised EBK sherds, while N-technique, thin walls (<10 mm) and a fine temper grain size were common in TRB sherds. Many of the samples did not fall neatly into either category and were assigned a 'probable' cultural association. To improve the contextualisation of the samples, we obtained radiocarbon dates for ten foodcrusts, five EBK and five TRB.

Previous research at the site extracted DNA from a piece of birch tar chewing gum dated to 3700 cal. BC. The chewing gum contained *Anas platyrhynchos* (mallard duck), *Corylus avellana* (hazelnut), *Betula pendula* (silver birch) and human DNA. The *A. platyrhynchos* and *C. avellana* DNA are thought to derive from consumed foods, while the *B. pendula* is from the chewing gum itself (Jensen et al. 2019). The extracted human DNA contained no trace of Neolithic farmer ancestry in the individual's genome (Jensen et al. 2019). This is a stark contrast to other studies that found a steep decline in the percentage of Western Hunter Gatherer (WGH) ancestry in favour of Neolithic farmers in samples from Danish Early Neolithic sites (Allentoft et al. 2024). Allentoft et al. (2024) also suggested that the remaining component of WGH ancestry in Neolithic individuals may actually derive from WGH-Neolithic admixture from before the Neolithic farmers' arrival in Denmark. Syltholm II

also deviates from the standard Neolithisation narrative in other ways. In the region generally, the start of the Neolithic also marks a dietary shift away from marine or freshwater resources and with an increased focus on dairy. However, at Syltholm II evidence for marine resource exploitation persists into the Neolithic (Cubas et al. 2020; Jensen and Sørensen 2023). These previous studies suggest that the story of the Mesolithic-Neolithic transition might have played out differently at Syltholm II than in many other parts of Denmark and even Europe.

We focus on the analysis of foodcrusts for three main reasons. First of all, foodcrusts can provide direct evidence about what resources were cooked, making it very useful to study the introduction of domesticates. Secondly, foodcrusts are able to provide information on both the animal and plant component of the diet. Lastly, as there is no clear stratigraphy, pottery sherds are one of the few types of artefacts that can be clearly associated with a particular culture, while the organic material of the foodcrusts allows this to be absolutely dated as well.

The most commonly used method for the study of foodcrusts is lipid analysis, but lipid analysis is biased towards the detection of lipid-rich foods and is limited in its ability to distinguish the individual ingredients from a mixture. To circumvent these limitations we apply three different methods: Scanning Electron Microscopy (SEM), lipid analysis and protein analysis. By combining methods that favour plant remains, lipid-rich foods and protein-rich foods a more comprehensive understanding of the foods cooked in the vessels will emerge.

5.2 Results

In this paper we present the results of 76 foodcrusts, of which 10 were directly dated using ^{14}C dating, 20 were analysed with lipid analysis, 31 with SEM analysis and 73 were analysed proteomically (Fig. 1).

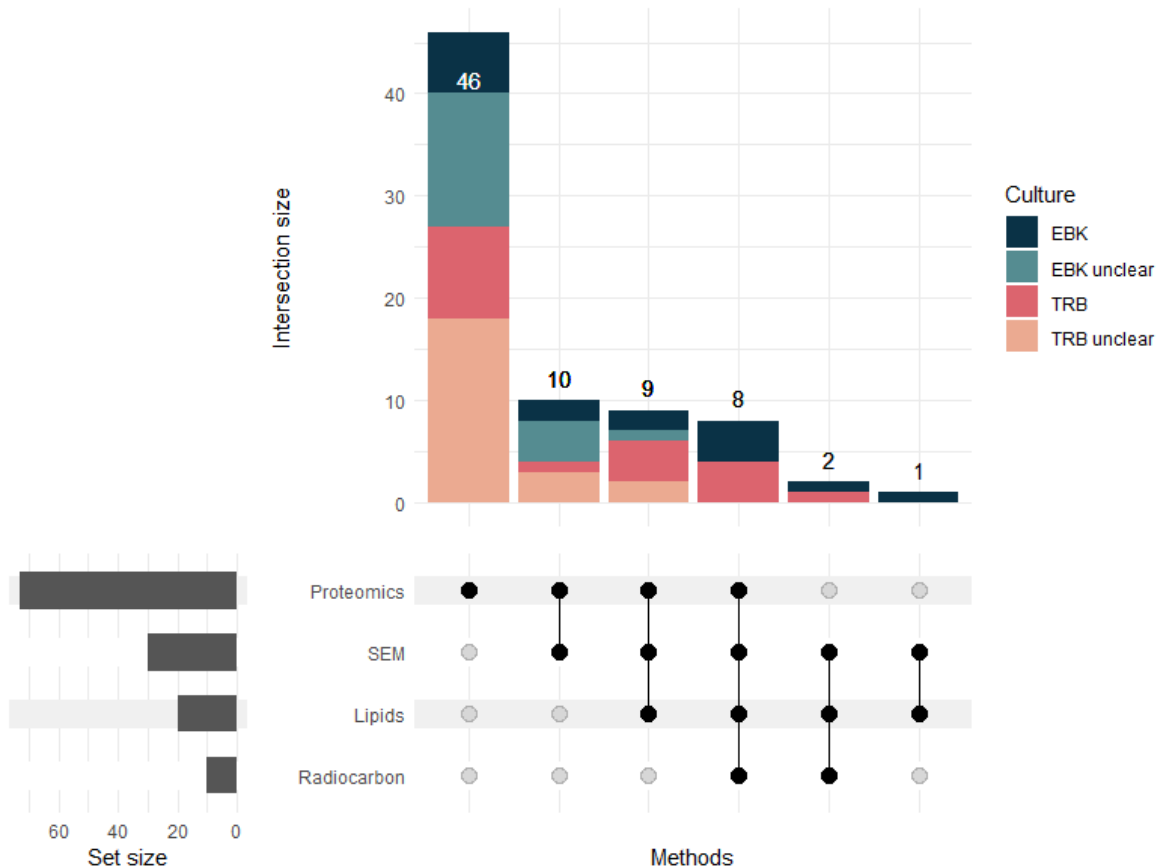


Figure 1, Number of samples and their cultural associations per method.

5.2.1 Radiocarbon dating

For five foodcrust that were identified as ‘clear’ EBK and five that were identified as ‘clear’ TRB we obtained radiocarbon dates calibrated using OxCal v4.4.4 (Bronk Ramsey 2020) (Fig 2. and SI 1). No clear temporal distinction is visible between the two cultures. A marine reservoir effect (MRE) is to be expected for these samples. A recent study showed that for samples, such as foodcrust, that could contain marine resources from the Baltic as well as local aquatic resources out the fjord a reservoir effect of around 400 years should be expected (Philippsen 2018). However, as the average $\delta^{13}\text{C}$ of the dated samples is -27.3 ± 1.0 ‰, while for marine-dominated samples a value of -10 ‰ is expected, it is assumed that the effect of the reservoir effect on the dates is minor. It would likely also not change the distribution of dates, assuming the EBK samples contain an equal or larger amount of marine foodstuffs.

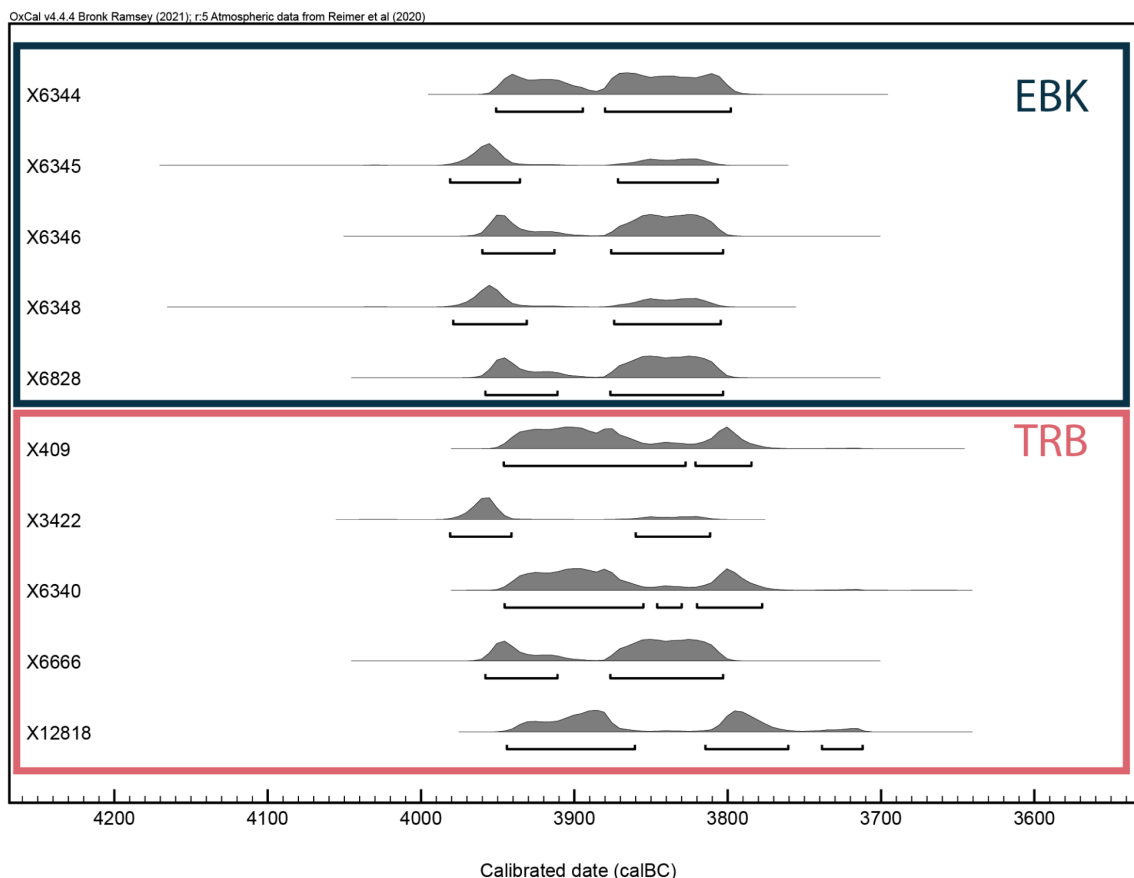


Figure 2, 95% confidence intervals of calibrated ^{14}C dates of five EBK and 5 TRB foodcrusts using OxCal v4.4.4 (Bronk Ramsey 2020).

5.2.2 SEM

EBK

Microremains revealed by SEM represented a wide range of taxa, although not all foodcrusts investigated contained diagnostic microremains. In two of EBK foodcrusts we observed what resembles the epidermis of a berry cf. *Viburnum opulus*. The same foodcrusts (Fig. 3A-B) also show objects that have been previously interpreted as sturgeon bone (Bondetti et al. 2021). Another EBK sample featured numerous instances of tuber parenchyma (Fig. 3C), which could potentially be attributed to *Beta vulgaris maritima* (sea beet), although this identification is not certain. Two other foodcrusts show seeds likely from the same family as *B. vulgaris maritima*, Chenopodiaceae, but one of these is fairly unspecific, while the other has seeds that resemble cf. *Atriplex* spp. (orache). The foodcrust from the sherd of a probable EBK vessel contained an object (Fig. 3D) that resembled the aleurone layer of a cell cf. *Hordeum vulgare* in a way similar to ancient bread remains (Arranz-Otaegui et al. 2018).

TRB

Only four of the TRB foodcrust contained diagnostic fragments. One of these showed damaged tuber parenchyma and seeds, but also cereal bran cf. *Hordeum vulgare*. More tuber parenchyma and seeds cf. *Atriplex* sp. were found in another TRB foodcrust. One TRB

foodcrust also featured remains (Fig. 3E) similar to the sturgeon bone and fish scales (Bondetti et al. 2021) that were observed in some of the EBK foodcrusts.

In one of the probable TRB foodcrusts evidence cf. *Viburnum opulus* was observed once more. Lastly, one of the probable TRB foodcrusts contained cereal bran, which based on the cell size is likely to be cf. *Triticum* sp. (Fig. 3F). The same sample also contains seemingly smooth oblong objects of the size of cells, which have not yet been securely identified.

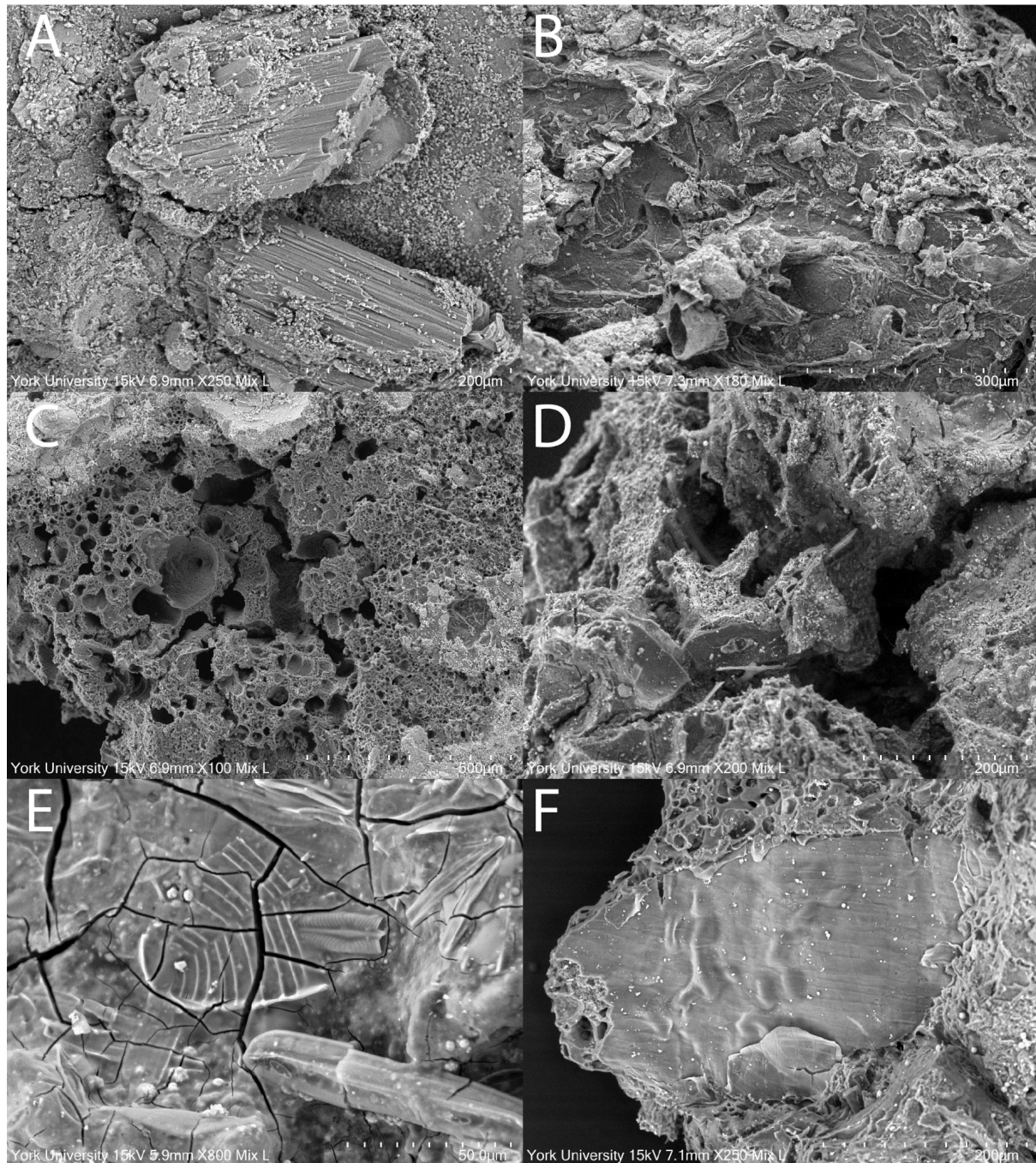


Fig 3, SEM images of foodcrust. **A**, sample X7340, fragment similar to sturgeon bone. **B**, sample X7340, epidermis cf. *Viburnum opulus*. **C**, sample X6828, cf. *Beta vulgaris*. **D**, sample X7281, aleurone layer cf. *Hordeum vulgare*. **E**, sample X4536, object resembling fish scales. **F**, sample X12455, cereal bran cf. *Triticum* sp..

5.2.3 Lipid analysis

In general the $\delta^{13}\text{C}_{16:0}$ and $\delta^{13}\text{C}_{18:0}$ profiles of the EBK and TRB samples analysed in this study look remarkably similar (Fig. 4). There are vessels from both cultures that fall clearly in the range of the marine resources and both cultures have samples with an unambiguous dairy signal. Additionally, a majority of the foodcrusts, including those who do not fall in the marine $\delta^{13}\text{C}_{16:0}$ and $\delta^{13}\text{C}_{18:0}$ range display marine biomarkers. Both cultures also contain samples of which the $\delta^{13}\text{C}_{16:0}$ and $\delta^{13}\text{C}_{18:0}$ ratios suggest primarily a ruminant adipose fat origin. The presence of ruminant adipose TAGs in 3 EBK and 2 TRB samples corroborates this. A fair number of samples does not fall neatly within any of the displayed reference ranges. It is likely that these samples are a mixture of foodstuffs with different isotopic values.

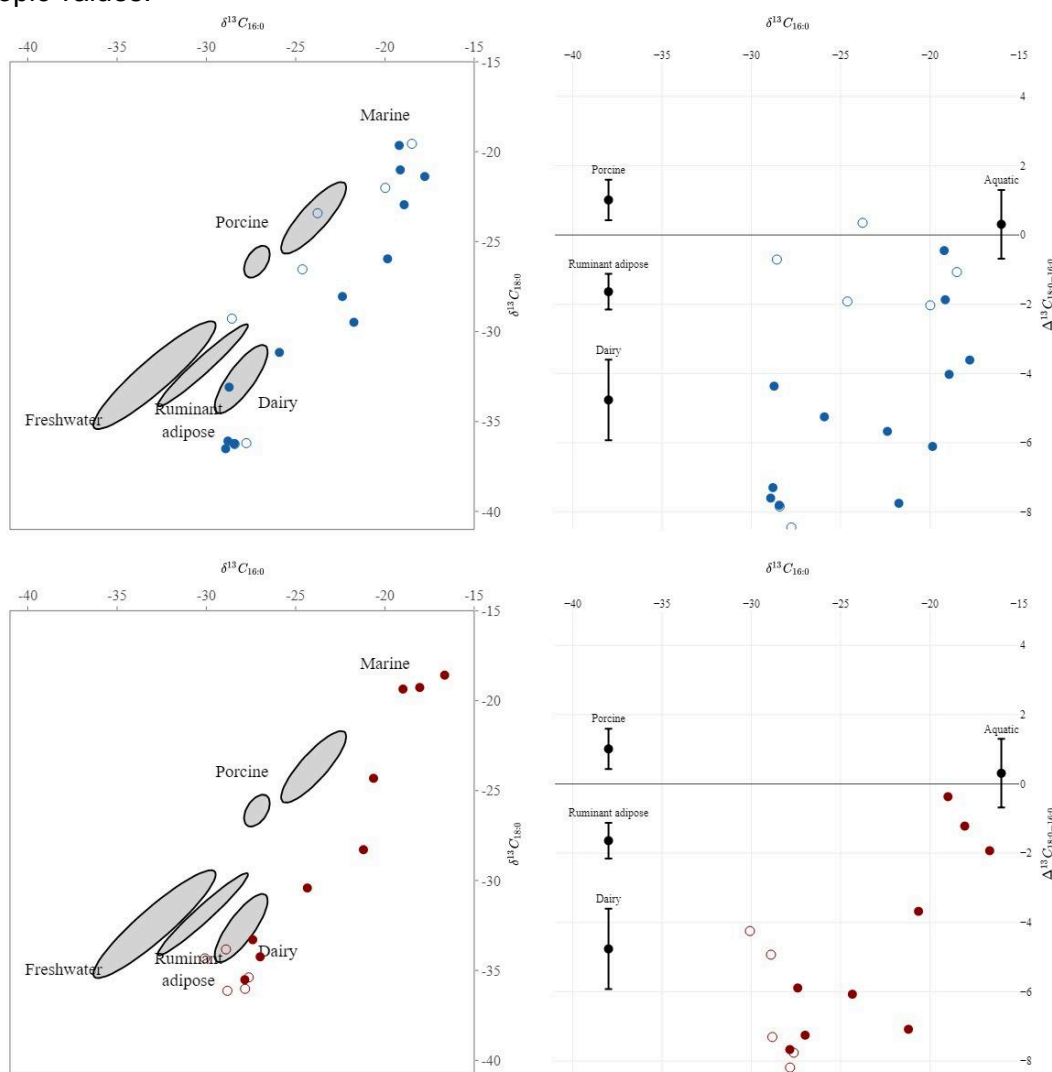


Figure 4, Compound-specific isotope ratios. EBK foodcrusts are displayed top in blue, TRB foodcrusts bottom in red.

5.2.4 Proteomics

49 foodcrusts contained a total of 524 distinct peptide sequences following initial steps of filtering out all peptides identified in the blank samples, presumed contaminants (e.g. human, mice and rat proteins) and all proteins with less than two associated peptides

(SI 2). Unfortunately, the Unipept (Gurdeep Singh et al. 2019) analysis revealed that the majority of these sequences were shared at a very high taxonomic level. Unipept could only assign a taxonomic identification at order level or lower to 107 sequences. After a second step of taxonomic verification using BLAST (Altschul et al. 1990) and refiltering the data to require two peptide sequences per taxon and protein, 66 peptide sequences remain between 20 proteins (Fig. 5A). This final set of peptide identifications can be divided into three groups. The first group of peptides are alpha-s1-casein and beta-lactoglobulin peptides assigned to Bovinae and Pecora. The second group consists of peptides assigned to vicilin, convicilin and provicilin (type A/B), which are all seed storage proteins. These proteins are associated with *Pisum sativum* (garden pea), Fabaeae, the IRL clade and Eukaryota. The last group only consists of two *Xenopus laevis* (African clawed frog) peptides of a cytoskeleton associated protein (Q9PT63).

During the taxonomic verification of the observed peptide sequences no prior assumptions were made regarding the potential taxa. However, especially in the case of dairy proteins, the archaeological context can drastically reduce the number of potential species. In the case of Lolland during the Mesolithic-Neolithic transition, three species could be expected: *Bos taurus* (cattle), *Capra hircus* (goat) and *Ovis aries* (sheep). All three are domesticated and *Bos primigenius* (wild auroch) is extinct on Lolland by this time (Bennike and Jessen 2023). To determine which of these three species likely contributed the dairy proteins in the foodcrust, their reference sequences were aligned to the observed peptide sequences (Fig. 5B, 3C). Two of the SAPs (single amino acid polymorphisms) between *B. taurus* and *C. hircus* and *O. aries* on alpha-s1-casein and one on beta-lactoglobulin are potential deamidation sites and should not be used for species identification. Nonetheless, for all SAPs on the observed peptides the sequence conforms to the *B. taurus* reference and no sequence in disagreement with a *B. taurus* source was observed.

As a last verification step we calculated the percentage of deamidated PSMs of the observed proteins, particularly the glutamine to glutamic acid deamidation. Due to the limited number of peptide residues it was not possible to calculate the deamidation rate for all proteins, but in general the dairy proteins showed high deamidation values (75-94%) with the exception of samples X3182, X3528, X5018 and X6583, of which the peptides were undeamidated. The plant proteins showed a much lower rate of deamidation, ranging between 3.8% and 7.5%. Lastly, the *Xenopus laevis* protein had a deamidation (Q) value of 50%.

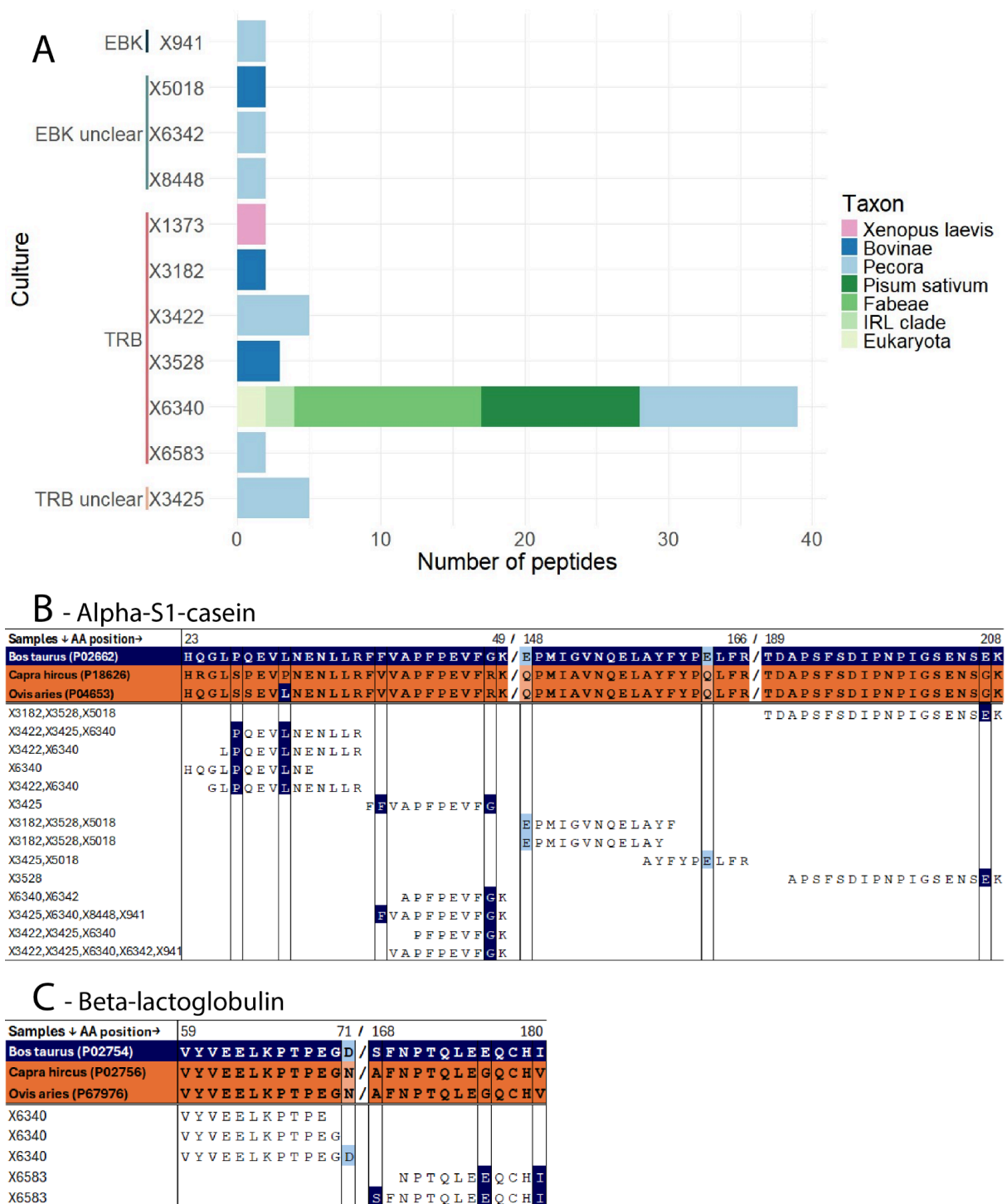


Figure 5, A, The number of peptides and associated taxa per sample after validation. All Bovinae and Pecora peptides are derived from alpha-S1-casein and beta-lactoglobulin. All Eukaryota, IRL clade, Fabeae and *Pisum sativum* peptides belong to the proteins vicilin, convicilin and provicilin (type A/B). **B,** All alpha-s1-casein peptides aligned to *Bos taurus* (P02662), *Capra hircus* (P18626) and *Ovis aries* (P04653) reference sequences. Sample sequences are coloured with the matching reference for SAPs, potential deamidation sites have a lighter colour. **C,** All beta-lactoglobulin peptides aligned to *Bos taurus* (P02754), *Capra hircus* (P02756) and *Ovis aries* (P67976) reference sequences. Sample sequences are coloured with the matching reference for SAPs, potential deamidation sites have a lighter colour.

5.3 Discussion

The results of the SEM, lipid and proteomic analysis of the Syltholm II foodcrusts differ substantially in the foodstuffs that they managed to identify. This lack of overlap emphasises how these methods can complement each other especially when applied to the same samples, but also highlights the risk of missing particular resources if these methods are applied in isolation. Before integrating our results and situating them in the context of the Mesolithic-Neolithic transition it is essential to establish the reliability and taxonomic precision of our data.

5.3.1 Reliability

Lipid and SEM analysis have been routinely applied to foodcrusts over the course of several decades (Heron and Evershed 1993; Craig 2004; Heron and Craig 2015; Lyu et al. 2024; Kubiak-Martens 1999; Kubiak-Martens et al. 2024) and the data presented in this chapter from these methods matches the established quality criteria (González Carretero, Wollstonecroft and Fuller 2017; Craig et al. 2012; Cramp and Evershed 2014), showing no reason to question their reliability. However, proteomic analysis of archaeological foodcrusts has been more rarely applied and a combination of some of the observed taxa with the low deamidation values warrants further discussion on the reliability of the identified proteins. Particularly suspicious identifications are *Xenopus laevis* and *Pisum sativum* along with associated higher taxonomic ranks. *Pisum sativum* has been found at TRB sites in Poland (Bieniek 2007), but neither *P. sativum* nor other legumes are commonly found at Danish TRB sites. Additionally, the observed deamidation values of the *P. sativum* proteins are much lower compared to what is commonly found in archaeological proteins (Ramsøe et al. 2020). However, there are no clear thresholds for deamidation and legume proteins in a Neolithic sample are not extraordinary. For example, legumes were identified in calcified deposits on pottery from Çatalhöyük, but the majority of the associated peptides were deamidated (Hendy et al. 2018a). Therefore caution in interpreting the *Pisum sativum* proteins presented here is warranted.

As for *Xenopus laevis*, proteins from this species have been reported earlier as false positives arising due to peptide sequences being shared across many species (Barker et al. 2012), but in this study the sequence specificity was confirmed using BLAST and the associated proteins' deamidation values (50% for Q) are closer to what would be expected from an ancient sample. Evidence for the consumption of frogs is rare, although there are several clear cases, such as the 865 *Rana temporaria* (common frog) bones that were found in hearths at the 3000-2800 BC Czech site Kutná Hora-Denemark (Kyselý 2008). Additional evidence comes from the Belgian Mesolithic Swifterbant site Doel. Hearths from sector M of this site, which was dated to 4368-4156 cal. BC (Boudin et al. 2010; Van Neer et al. 2013), contained a number of burnt *Anura* sp. (frog or toad) bones. Hearth 3 in particular contained a high number of *Rana temporaria* remains, perhaps indicative of consumption (Van Neer et al. 2013). Considering that the *X. laevis* peptides withstood all taxonomic validation tests, the consumption of amphibians is perhaps not unthinkable, although it is likely to be a different species than *X. laevis*.

5.3.2 Taxonomic accuracy

Assuming that all identified proteins are in fact ancient, the question remains whether they actually belong to the taxon assigned to them by the BLAST search. The *X. laevis* protein is a good example, as it is not native to Europe and is a model organism. The identified peptides were likely assigned to *X. laevis*, because their true match was not in the database, which frequently occurs for many wild species (e.g. *Capreolus capreolus* samples being identified as bovid proteins in Chapter 3). Determining what taxon is the most probable origin of a protein can be challenging. The gene encoding for the observed *X. laevis* protein has not been annotated in any other amphibian species and there are no other proteins with similar sequences (i.e. none with $\geq 90\%$ similarity) (Suzek et al. 2015). The *P. sativum* peptides share a similar issue. Other Fabaceae, such as *Vicia sativa*, have been identified before at sites of similar ages, albeit EBK rather than TRB (Kubiak-Martens 2002). The reference proteome of *V. sativa* is a fraction of the size of *P. sativum* (240 vs 134705 protein sequences (Schoch et al. 2020)), meaning that it is highly likely that proteins of *V. sativa* or other Fabaceae with poor reference proteomes would be identified as *P. sativum*. Consequently, it seems there is no clear way to account for the coverage of reference proteomes in assigning taxonomic identities. Instead, we propose that in such cases where a protein appears to be species or genus specific, but there is reasonable doubt that all closely related taxa can actually be excluded, to report the taxonomic identifications as “confirming to (cf)” the species, similarly to many other biological disciplines. By reporting our observed proteins as cf. *Xenopus laevis* and cf. *Pisum sativum* we maintain the necessary nuance, while also signalling that these proteins were designated as species-specific by the used validation tools, Unipept and BLAST in this case.

5.3.3 Combining the data

Combining the evidence from all three methods, we have found evidence for the cooking of tubers, cf. *Beta vulgaris*, berries cf. *Viburnum opulus*, as well as seeds from Chenopodiaceae such as cf. *Atriplex* spp., along with the use of marine resources, ruminant adipose fats and dairy in the EBK vessels. The bone fragments observed with the SEM are similar to *Acipenser* sp., which has been found previously at other EBK sites (Enghoff 2011), but the marine signal measured in the lipid analysis could derive from any combination of many different species. The most commonly found fish species at MLF00906-II are *Pleuronectes platessa*, *Platichthys flesus* and *Limanda limanda*, followed at some distance by *Anguilla anguilla* (Groß 2024, personal communication). The assemblage at MLF00906-II differs somewhat from the average composition of Danish EBK sites in that gadids are usually the most abundant taxon (Price et al. 2018; Enghoff, MacKenzie and Nielsen 2007; Astrup et al. 2021; Pickard and Bonsall 2007), while at MLF00906-II they are only represented by two remains. However, an abundance of flatfish is in line with general EBK exploitation patterns, as they are usually the second most abundant after gadids, although there are other sites where they are the most abundant fish species as well (Fischer et al. 2021). Unfortunately the protein analysis is not able to further specify the fish taxon, as no fish proteins remained after the validation process. The lack of fish proteins was surprising, as previous experiments suggested that fish proteins are more resistant to diagenesis than plant or meat proteins (Chapter 2). However, it must be noted that in a recent blind test of proteomic taxonomic identification of foodcrusts, no proteins from the order of flatfish

(Pleuronectiformes) could be identified, even though plaice was cooked in two samples (Chapter 3).

The TRB and EBK foodcrusts are fairly similar. They contain evidence of marine resources, ruminant adipose fats and dairy. The main difference between the two datasets lies in the observed plant remains, which for the TRB include cereals such as cf. *Hordeum vulgare*. Remains of cf. *Triticum* sp. were found in a likely TRB vessel, but its cultural association was ambiguous. Interestingly, the two TRB foodcrusts with *Hordeum vulgare* and tuber remains showed a strong dairy signal in the compound specific isotope ratios, while the TRB foodcrust with evidence of *Viburnum opulus* plotted on the other end towards a heavy marine signal. Mixing of *V. opulus* and fish has been suggested earlier for early ceramics at Zamostje 2 in Russia (Bondetti et al. 2020). The fish species used in the TRB foodcrusts is likely to be the same as the fish used in the EBK foodcrusts. In general there was no large change in which fish species were exploited during the Mesolithic-Neolithic transition (Enghoff 2011) and the faunal assemblage found at MLF00906-II cannot be split in a Mesolithic and Neolithic component.

The foodcrust containing cf. *Hordeum vulgare* was one of the samples selected for radiocarbon dating and its date 3959-3908 and 3878-3803 cal. BC (OxCal v4.4.2 (Bronk Ramsey 2020)) places it among the earliest remains of cereals in Denmark (Sørensen and Karg 2014).

5.3.4 Dairy in the Mesolithic

A previous study found that about 22% of the analysed EBK vessels had compound specific isotope ratios matching with a dairy signal (Papakosta, Oras and Isaksson 2019; Lucquin et al. 2023) and this study complements this by providing further lipid isotope evidence along with dairy proteins. Considering the lack of evidence for animal husbandry it has been suggested that the dairy found in EBK foodcrusts must derive from trade with neighbouring TRB settlements (Lucquin et al. 2023). If dairy products were traded between these two cultures this raises the question of the form in which it was traded. Although BLG (beta-lactoglobulin) has been frequently observed in dental calculus and is considered to be one of the best preserving proteins (Hendy et al. 2018b; Fonseca, Freeman and Collins 2024), it was only identified in one (TRB) sample. The other, more frequently observed dairy protein in our samples is alpha-s1-casein. The small number of protein-yielding samples warrants caution, but as BLG is described as a biomarker of milk (Warinner et al. 2014), as opposed to hard cheese, its absence might suggest that other dairy products such as cheese or butter, which are relatively enriched in casein (Charlton et al. 2019; Greco et al. 2018; Evans et al. 2023), are more probable sources.

5.3.5 Change and continuity

Traditionally, the Mesolithic-Neolithic transition has been interpreted as a drastic change affecting the environment, economy and genetic ancestry and this seems to apply to Denmark as well (Allentoft et al. 2024). However, the archaeological record at Syltholm II suggests continuity rather than change throughout this period, including in genetic ancestry (Jensen et al. 2019). Remains of domestic mammals have been found at Syltholm II and they make up ~21% of the number of identified remains (Gron et al. 2024). Without stratigraphic evidence or direct dates to the contrary the domestic mammals are presumed

to correspond to the Neolithic occupation of the site (Jensen and Sørensen 2023), which would indicate a shift in exploited species from the Mesolithic to the Neolithic.

However, the only clear difference in the cooked foods, as indicated by SEM, lipids and proteomic analysis of foodcrusts, is the introduction of cereals in the TRB vessels. The earliest undisputed evidence of cereals in Denmark comes from radiocarbon dates of charred grains of emmer, einkorn, bread wheat and naked barley dating between 4000 and 3700 BC (Sørensen and Karg 2014). During this period other evidence of agriculture in the form of plough marks (Beck 2009), cereal pollen (Andersen 1990) and threshing waste (Skousen 2008) also starts to appear in Denmark.

The consumption of all other observed resources, including wild plants, marine foods and dairy products does not appear to differ between the EBK and TRB foodcrusts. Furthermore, direct radiocarbon dates of the foodcrusts of shards typologically and technologically identified as clearly TRB or EBK are not able to split the two types into distinct phases. From a pure theoretical perspective, the radiocarbon dates provide equal support to the hypothesis that the EBK predates the TRB, succeeds the TRB or is contemporary to the TRB. We do not claim that the EBK and TRB are contemporary at Syltholm II, but the lack of a clear EBK-TRB separation highlights the narrow timespan within which the analysed samples were deposited.

The lack of change could be interpreted as the result of a local population of hunter-gatherers remaining in place and adapting to new material culture, but this possibility has been rejected by many previous scholars (Lucquin et al. 2023). Alternatively, the short timespan and lack of a clear chronological EBK-TRB distinction could mean that the studied assemblage represents the transition in progress, similar to a “Negotiation phase” as suggested previously (Gron and Sørensen 2018). Regardless of how the EBK-TRB transition precisely occurred, it is clear that through the first centuries of the 4th millennium the inhabitants of Syltholm II exploited wild and domestic resources and while their ceramic technology changed, a component of hunter-gatherer ancestry remained in the population at least until around 3700 cal. BC.

5.4 Materials and methods

In total we analysed foodcrust samples from 76 sherds from MLF00906-I/II. Of these 16 had been classified as Ertebølle vessels based on their typology and technological aspects, such as coiling technique. 19 others had been classified as TRB. For the remaining 39 foodcrusts the cultural association was slightly ambiguous. 18 sherds were considered to likely represent EBK vessels, while the remaining 23 sherds are thought to derive from TRB pots. Considering the lack of clear stratigraphical layers and in some cases the ambiguity of the typological and technological analysis, there was some concern regarding the security of the context. In order to address this, samples from 10 foodcrusts (5 EBK, 5 TRB), were sent to NTNU for radiocarbon analysis.

Not all sherds contained sufficient foodcrust to perform SEM, lipid and protein analysis. We attempted to apply as many different methods on the same sample as possible and an overview of all analyses applied per vessel can be found in SI 3.

5.4.1 Ceramic typo-technological analysis

The cultural association of the analysed sherds was primarily based on the coiling technique, wall thickness, surface treatment, rim shape, rim orientation, belly or neck shape,

bottom shape, decoration and temper material and size. In general, U-technique coiling, thick walls (>10 mm), granite/feldspar temper with a medium grain size (1-5 mm) were associated with the EBK, while N-technique coiled, thin-walled (<10 mm) and fine grained tempered (grain size < 1 mm) shards were assigned as TRB. The temper of the TRB sherds varied between quartz, sand and mica. None of these criteria was decisive on its own and some EBK-identified vessels were fine grained or thin-walled, while other TRB shards were medium grained or thick walled. However, no TRB shard was medium grained **and** thick-walled and U-technique coiled shards were only assigned to the EBK, probable EBK or as indeterminate.

5.4.2 Radiocarbon analysis

Radiocarbon analysis was performed at the National Laboratory of Age Determination at the Norwegian University of Science and Technology in Trondheim, Norway. In short, samples were treated with an acid-alkali-acid preparation and then reduced to produce graphite and ^{14}C age measurements were performed using an Accelerator Mass Spectrometer (AMS) (Seiler et al. 2019).

5.4.3 Foodcrust microscopic analysis

An initial screening was performed using a low-powered Leica MZ APO binocular microscope at magnifications between $8\times$ - $50\times$. 10 EBK, 5 probable EBK, 11 TRB and 5 probable TRB foodcrusts were marked as promising and were subjected to SEM analysis using a Hitachi TM4000 at the University of York. Guidelines for the quantification, measurement and categorisation of plant and animal inclusions and voids from González Carretero *et al.* (González Carretero, Wollstonecroft and Fuller 2017) were followed.

5.4.4 Lipid analysis

The lipid analysis was performed according to Lucquin et al. (2023) and focused on 8 EBK, 1 probable EBK, 9 TRB and 2 probable TRB samples. In brief, a Total Lipid Extract (TLE) protocol was performed using ~20 mg of foodcrust, which were acidified with concentrated sulphuric acid, heated for 4 hours at 70°C and finally extracted with phase separation under nitrogen flow using *n*-hexane. The lipid extracts were first analysed with GC-MS and compound specific isotope ratios of palmitic and stearic acid were obtained with GC-C-IRMS.

5.4.5 Proteomic analysis

Protein extraction

Protein extraction from 14 EBK, 18 probable EBK, 18 TRB and 23 probable TRB foodcrusts, including from the interior and exterior, was performed using a DES-based (Deep Eutectic Solvent) protocol (Pal Chowdhury et al. 2022), as described elsewhere (Dekker and Ramsøe 2025). In short, ~20 mg of foodcrust was incubated in 300 μL of a 2:1 molar ratio urea:GuHCl (guanidine hydrochloride) solution at 65°C for four hours. After incubation the samples were removed from the heating block and stabilised by adding 0.5 mL purified water and centrifuged at 15000 rcf for 15 minutes. Reduction and alkylation was achieved by adding 15 μL of a 100 mM TCEP (Tris(2-carboxyethyl)phosphine) and CAA

(2-Chloroacetamide) solution, homogenising, centrifuging at 15000 rcf for one minute and incubating at 99°C for 10 minutes. An initial cleaning step was performed using C18 StageTips. The StageTips were conditioned with 150 µL 100% methanol, washed with 150 µL 80% ACN (acetonitrile) + 0.1% TFA (trifluoroacetic acid) and a second wash of 150 µL 0.1% TFA. Following these washes 150 µL of sample was loaded on the tips and the tips were washed twice with 150 µL of 0.1% TFA. The proteins were then eluted off the tips using 50 µL 50% ACN + 0.1% TFA and the solution was dried down to near completion in a speedvac at 60°C in order to be resolubilised in 100 µL 50 mM ABC (ammonium bicarbonate).

The extracted proteins were first digested by adding 1 µL of a 0.4 µg/µL chymotrypsin (Promega) solution to each sample and incubating at 25°C for 18 hours. Secondly, 1 µL of 0.4 µg/µL trypsin (Promega) was added to each sample and incubated at 37°C for 18 hours. The digested samples were acidified by adding 10 µL 5% (v/v) TFA and subjected to a final cleaning step using EvoTips (EvoSep) following a slight modification of the manufacturer's protocol. For each sample, two EvoTips were prepared, to be analysed using different mass spectrometry settings. The EvoTips were conditioned by adding 20 µL of 0.1% formic acid (FA) in ACN and then submerged in isopropanol for one minute. After removing the EvoTips from the isopropanol they were washed with 20 µL of 0.1% FA, after which 20 µL sample was loaded on each tip. The tips were then cleaned thrice with 50 µL 0.1% FA. 100 µL 0.1% FA was loaded on top of each tip to ensure they stayed wet before mass spectrometric analysis.

Mass spectrometry data acquisition

The extracted peptides were loaded on an EvoSep One nanoflow UPLC (EV-1000, EvoSep) using the 100 SPD (samples per day) gradient (from EvoSep One HyStar Driver 2.4.125.0) on an EvoSep C18 Performance column (EV1109, 8 cm x 150 µm, 1.5 µm). The EvoSep was coupled to a timsTOF HT mass spectrometer (Bruker) with a CaptiveSpray ionisation source. Positive PASEF-DDA, ESI-MS and MS2 spectra were acquired using Compass HyStar software (version 6.2, Bruker). The instrument settings were as follows; instrument source: capillary voltage, 1,600 V; dry gas, 3 l/min; dry temperature; 180°C. Mass range: m/z 100-1,700. TIMS settings were: 1/K0 0.6-1.60 V.s/cm²; Ramp time, 100 ms; Ramp rate 9.42 Hz. Data dependent acquisition was performed with 10 PASEF ramps and a total cycle time of 1.17 s. An intensity threshold of 2,500 and a target intensity of 20,000 were set with active exclusion applied for 0.4 min post precursor selection. CID was used to fragment the precursor ions and collision energy was interpolated between 20 eV at 0.6 V.s/cm² to 59 eV at 1.6 V.s/cm².

Bioinformatic analysis

A two-step analysis approach using Fragpipe (v21) (Kong et al. 2017; Yu et al. 2020) was adopted. It consisted of an initial open search with trypsin and chymotrypsin as enzymes against the entire UniProt database, followed by a second closed search. The fasta file of identified protein sequences automatically output by the open search was then used as database for the closed search, which was done in semi-specific (chymo)trypsin cleavage mode. The results of the closed search were used for further downstream analysis.

A minimum of two peptides per protein were required for accepting a protein's presence and all peptides identified in any of the blank samples were excluded from all samples to eliminate any form of lab or carry over contamination. Additionally, all keratins,

proteases and proteins assigned to humans, rats or mice were excluded as well. Only peptides that were marked as being unique to a protein were used in the taxonomic identification process.

The remaining peptide sequences were uploaded to the Unipept web service (v5.1.5) (Gurdeep Singh et al. 2019) for an initial taxonomic assignment. Unipept was run with advanced missed cleavage handling turned on and leucine and isoleucine amino acids were considered interchangeable in its matching to UniProt protein sequences. Of the resulting peptide-taxon matches peptides belonging to the phylum Chordata or Streptophyta and whose taxonomic assignment was to order level or more specific were submitted to BLAST (Altschul et al. 1990) for a second step of taxonomic verification. The MRCA of all species with a 100% query similarity and a score difference <5 from the top hit of BLAST was annotated as the peptide's MRCA.

In order to refine the taxonomic identification of the dairy proteins, all peptides associated with alpha-s1-casein and beta-lactoglobulin were aligned using Clustal Omega (Sievers et al. 2011; Sievers and Higgins 2018; Sievers, Barton and Higgins 2020) on the UniProt server with the species that, based on the archaeological context, might have been present (i.e. *Bos taurus*, *Capra hircus* and *Ovis aries*).

Deamidation was calculated using a pilot version of the Python script Anubis (Chiang 2024). Anubis calculates site-specific deamidation rates and provides a protein-level summary.

References

- Allentoft, M. E. et al. (2024). 100 ancient genomes show repeated population turnovers in Neolithic Denmark. *Nature*, 625 (7994), pp.329–337. [Online]. Available at: doi:10.1038/s41586-023-06862-3.
- Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410. [Online]. Available at: doi:10.1016/S0022-2836(05)80360-2.
- Andersen, S. T. H. (1990). Pollen spectra from two early neolithic lugged jars in the long barrow at bjørnsholm, Denmark. *Journal of Danish archaeology*, 9 (1), Det Kgl. Bibliotek/Royal Danish Library., pp.59–63. [Online]. Available at: doi:10.1080/0108464x.1990.10590035.
- Arranz-Otaegui, A. et al. (2018). Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *Proceedings of the National Academy of Sciences*, 115 (31), National Acad Sciences., pp.7925–7930. [Online]. Available at: doi:10.1073/pnas.1801071115.
- Astrup, P. M. et al. (2021). A drowned Mesolithic shell midden complex at Hjarnø Vesterhoved, Denmark and its wider significance. *Quaternary science reviews*, 258, p.106854. [Online]. Available at: doi:10.1016/j.quascirev.2021.106854.
- Barker, A. et al. (2012). An Optimized Approach for Protein Residue Extraction and Identification from Ceramics After Cooking. *Journal of Archaeological Method and Theory*, 19 (3), pp.407–439. [Online]. Available at: doi:10.1007/s10816-011-9120-5.
- Beck, M. R. (2009). Lå Danmarks første pløjemark ved Egense. *Svendborgs Museums Årbog*, 7e16.
- Bennike, O. and Jessen, C. (2023). Environmental changes after the last deglaciation, southern Lolland, Denmark. In: Groß, D. and Rothstein, M. (Eds). *Changing identity in a changing world*. Leiden, Netherlands : Sidestone Press. pp.33–42. [Online]. Available at: doi:10.59641/pfm7c6gh.
- Bieniek, A. (2007). Neolithic plant husbandry in the Kujawy region of central Poland. In: *The Origins and Spread of Domestic Plants in Southwest Asia and Europe*. 1st Edition. Routledge. pp.343–358. [Online]. Available at: doi:10.4324/9781315417615-25 [Accessed 6 October 2024].
- Bondetti, M. et al. (2020). Fruits, fish and the introduction of pottery in the Eastern European plain: Lipid residue analysis of ceramic vessels from Zamostje 2. *Quaternary international: the journal of the International Union for Quaternary Research*, 541, pp.104–114. [Online]. Available at: doi:10.1016/j.quaint.2019.05.008.
- Bondetti, M. et al. (2021). Neolithic farmers or Neolithic foragers? Organic residue analysis of early pottery from Rakushechny Yar on the Lower Don (Russia). *Archaeological and anthropological sciences*, 13 (8), p.141. [Online]. Available at: doi:10.1007/s12520-021-01412-2.
- Boudin, M. et al. (2010). Fish reservoir effect on charred food residue 14C dates: Are stable isotope analyses the solution? *Radiocarbon*, 52 (2), Cambridge University Press (CUP)., pp.697–705. [Online]. Available at: doi:10.1017/s0033822200045719 [Accessed 29 October 2024].
- Bronk Ramsey, C. (2020). *OxCal*. Oxford : University of Oxford. [Online]. Available at:

<https://c14.arch.ox.ac.uk/oxcal.html>.

Charlton, S. et al. (2019). New insights into Neolithic milk consumption through proteomic analysis of dental calculus. *Archaeological and anthropological sciences*, 11 (11), Springer., pp.6183–6196. [Online]. Available at: doi:10.1007/s12520-019-00911-7.

Chiang, Y. (2024). *Anubis*. Github. [Online]. Available at: https://github.com/yc386/anubis_palaeoproteomics/blob/main/anubis_protein_lv.ipynb [Accessed 2 November 2024].

Courel, B. et al. (2020). Organic residue analysis shows sub-regional patterns in the use of pottery by Northern European hunter–gatherers. *Royal Society Open Science*, 7 (4), Royal Society., p.192016. [Online]. Available at: doi:10.1098/rsos.192016.

Craig, O. et al. (2012). Distinguishing wild ruminant lipids by gas chromatography/combustion/isotope ratio mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 26 (19), pp.2359–2364. [Online]. Available at: doi:10.1002/rcm.6349.

Craig, O. E. (2004). *Organic analysis of «food crusts» from sites in the Schelde valley, Belgium: a preliminary evaluation*. [Online]. Available at: https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=izEutEoAAAAJ:4TOpqqG69KYC [Accessed 18 October 2024].

Cramp, L. and Evershed, R. (2014). Reconstructing aquatic resource exploitation in human prehistory using lipid biomarkers and stable isotopes. *Treatise on geochemistry: archaeology and anthropology*, 14, Amsterdam: Elsevier., pp.319–339. [Online]. Available at: doi:10.1016/B978-0-08-095975-7.01225-0.

Cubas, M. et al. (2020). Latitudinal gradient in dairy production with the introduction of farming in Atlantic Europe. *Nature communications*, 11 (1), p.2036. [Online]. Available at: doi:10.1038/s41467-020-15907-4.

Dekker, J. and Ramsøe, M. (2025). *DES foodcrust protein extraction protocol*. [Online]. Available at: doi:10.17504/protocols.io.j8nlk9r6wv5r/v1 [Accessed 17 February 2025].

Enghoff, I. (2011). *Regionality and biotope exploitation in Danish Ertebølle and adjoining periods*, Scientia Danica. Series B Biologica. Det Kongelige Danske Videnskabernes Selskab.

Enghoff, I. B., MacKenzie, B. R. and Nielsen, E. E. (2007). The Danish fish fauna during the warm Atlantic period (ca. 7000–3900bc): Forerunner of future changes? *Fisheries research*, 87 (2-3), Elsevier BV., pp.167–180. [Online]. Available at: doi:10.1016/j.fishres.2007.03.004.

Evans, M. et al. (2023). Detection of dairy products from multiple taxa in Late Neolithic pottery from Poland: an integrated biomolecular approach. *Royal Society open science*, 10 (3), p.230124. [Online]. Available at: doi:10.1098/rsos.230124.

Fischer, A. et al. (2021). The Rødhals kitchen midden – marine adaptations at the end of the Mesolithic world. *Journal of archaeological science, reports*, 39 (103102), Elsevier BV., p.103102. [Online]. Available at: doi:10.1016/j.jasrep.2021.103102.

Fonseca, B., Freeman, C. L. and Collins, M. J. (2024). The survival of β -lactoglobulin peptides in the archaeological record: Vulnerability vs. Sequence variation. *bioRxiv*, p.2024.09.13.612646. [Online]. Available at: doi:10.1101/2024.09.13.612646 [Accessed 17 September 2024].

- González Carretero, L., Wollstonecroft, M. and Fuller, D. Q. (2017). A methodological approach to the study of archaeological cereal meals: a case study at Çatalhöyük East (Turkey). *Vegetation history and archaeobotany*, 26 (4), Springer Science and Business Media LLC., pp.415–432. [Online]. Available at: doi:10.1007/s00334-017-0602-6.
- Greco, E. et al. (2018). Proteomic Analyses on an Ancient Egyptian Cheese and Biomolecular Evidence of Brucellosis. *Analytical chemistry*, 90 (16), pp.9673–9676. [Online]. Available at: doi:10.1021/acs.analchem.8b02535.
- Gron, K. J. et al. (2024). Neolithisation through bone: Stable isotope analysis of human and faunal remains from Syltholm II, Lolland, Denmark. *Journal of Archaeological Science: Reports*, 53, p.104384. [Online]. Available at: doi:10.1016/j.jasrep.2024.104384.
- Gron, K. J. and Sørensen, L. (2018). Cultural and economic negotiation: a new perspective on the Neolithic Transition of Southern Scandinavia. *Antiquity*, 92 (364), Antiquity Publications., pp.958–974. [Online]. Available at: doi:10.15184/aqy.2018.71.
- Gurdeep Singh, R. et al. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of proteome research*, 18 (2), pp.606–615. [Online]. Available at: doi:10.1021/acs.jproteome.8b00716.
- Hartz, S., Lübke, H. and Terberger, T. (2007). From fish and seal to sheep and cattle: new research into the process of neolithisation in northern Germany. In: *Going Over: The Mesolithic-Neolithic Transition in North-West Europe*. British Academy. pp.567–594. [Online]. Available at: doi:10.5871/bacad/9780197264140.003.0027 [Accessed 13 September 2024].
- Hendy, J. et al. (2018a). Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nature communications*, 9 (1), p.4064. [Online]. Available at: doi:10.1038/s41467-018-06335-6.
- Hendy, J. et al. (2018b). Proteomic evidence of dietary sources in ancient dental calculus. *Proceedings. Biological sciences / The Royal Society*, 285 (1883), royalsocietypublishing.org. [Online]. Available at: doi:10.1098/rspb.2018.0977.
- Heron, C. and Craig, O. (2015). Aquatic Resources in Foodcrusts: Identification and Implication. *Radiocarbon*, 57 (4), Cambridge University Press., pp.707–719. [Online]. Available at: doi:10.2458/azu_rc.57.18454 [Accessed 12 May 2022].
- Heron, C. and Evershed, R. P. (1993). The analysis of organic residues and the study of pottery use. *Archaeological method and theory*, JSTOR. [Online]. Available at: https://www.jstor.org/stable/20170233?casa_token=yTiAlQm80ggAAAAA:tsY_ezNzdKT2Wp0-Zt_n6hxSMVa99awRsAYX2DaoOvxPZqPs_ohvhF_82pQdsFPQdFjm2n0FÖsdFGvbM_dgouuCCHHhgi10Re8lIn_Zg_gXxmWDSZ-H71A.
- Jensen, T. Z. T. et al. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nature communications*, 10 (1), p.5520. [Online]. Available at: doi:10.1038/s41467-019-13549-9.
- Jensen, T. Z. T. and Sørensen, L. V. (2023). Duality in the Early Neolithic on Lolland-Falster and in south Scandinavia. In: Groß, D. and Rothstein, M. (Eds). *Changing identity in a changing world*. Leiden, Netherlands : Sidestone Press. pp.89–107. [Online]. Available at: doi:10.59641/pfm7c6gh.
- Klassen, L. (2004). *Jade und Kupfer. Untersuchungen zum Neolithisierungsprozess im westlichen Ostseeraum unter besonderer Berücksichtigung der Kulturentwicklung Europas 5500-3500 BC*. Århus, Denmark : Aarhus University Press. [Online]. Available at:

doi:10.3828/BFARM.2005.4.3.

Kong, A. T. et al. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature methods*, 14 (5), Springer Science and Business Media LLC., pp.513–520. [Online]. Available at: doi:10.1038/nmeth.4256.

Krause-Kyora, B. et al. (2013). Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nature communications*, 4 (1), Springer Science and Business Media LLC., p.2348. [Online]. Available at: doi:10.1038/ncomms3348.

Kubiak-Martens, L. (1999). The plant food component of the diet at the late Mesolithic (Ertebolle) settlement at Tybrind Vig, Denmark. *Vegetation history and archaeobotany*, 8 (1-2), Springer Science and Business Media LLC., pp.117–127. [Online]. Available at: doi:10.1007/bf02042850.

Kubiak-Martens, L. (2002). New evidence for the use of root foods in pre-agrarian subsistence recovered from the late Mesolithic site at Halsskov, Denmark. *Vegetation history and archaeobotany*, 11 (1-2), Springer Science and Business Media LLC., pp.23–32. [Online]. Available at: doi:10.1007/s003340200003.

Kubiak-Martens, L. et al. (2024). Transformation of cereal grains: Botanical and chemical analysis of food residues encrusted on pottery from the Funnel Beaker settlement of Oldenburg LA 77, northern Germany. *PloS one*, 19 (1), p.e0296986. [Online]. Available at: doi:10.1371/journal.pone.0296986.

Kysely, R. (2008). Frogs as a part of the Eneolithic diet. Archaeozoological records from the Czech Republic (Kutná Hora-Denemark site, Řivnáč Culture). *Journal of archaeological science*, 35 (1), Elsevier BV., pp.143–157. [Online]. Available at: doi:10.1016/j.jas.2007.02.016 [Accessed 10 October 2024].

Lucquin, A. et al. (2023). The impact of farming on prehistoric culinary practices throughout Northern Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 120 (43), p.e2310138120. [Online]. Available at: doi:10.1073/pnas.2310138120.

Lyu, N. et al. (2024). Characterization of Pottery Foodcrusts Through Lipid and Proteomic Analyses: A Case Study from the Xiawan Site in Yixing City, East China. *Journal of archaeological science*, 161 (105902), papers.ssrn.com. [Online]. Available at: doi:10.1016/j.jas.2023.105902 [Accessed 8 September 2023].

Måge, B. T., Groß, D. and Kanstrup, M. (2023). The Femern project. In: Groß, D. and Rothstein, M. (Eds). *Changing identity in a changing world*. Leiden, Netherlands : Sidestone Press. pp.21–31. [Online]. Available at: doi:10.59641/pfm7c6gh.

Marchand, G. (2007). Neolithic fragrances: Mesolithic-Neolithic interactions in western France. *Proceedings-British Academy*, researchgate.net., pp.224–242. [Online]. Available at: doi:10.5871/BACAD/9780197264140.003.0012.

Pal Chowdhury, M. et al. (2022). Novel Deep Eutectic Solvent-Based Protein Extraction Method for Pottery Residues and Archeological Implications. *Journal of proteome research*, 21 (11), American Chemical Society., pp.2619–2634. [Online]. Available at: doi:10.1021/acs.jproteome.2c00340.

Papakosta, V., Oras, E. and Isaksson, S. (2019). Early pottery use across the Baltic—a comparative lipid residue study on Ertebølle and Narva ceramics from coastal hunter-gatherer sites in southern Scandinavia, northern Germany and Estonia. *Journal of archaeological science*, 24, Elsevier., pp.142–151. [Online]. Available at:

doi:10.1016/j.jasrep.2019.01.003.

Philippesen, B. (2018). Reservoir Effects in a Stone Age Fjord on Lolland, Denmark. *Radiocarbon*, 60 (2), Cambridge University Press., pp.653–665. [Online]. Available at: doi:10.1017/RDC.2018.6 [Accessed 26 October 2022].

Pickard, C. and Bonsall, C. (2007). Late Mesolithic coastal fishing practices. The evidence from Tybring Vig, Denmark. In: Hardh, B., Jennbert, K. and Olausson, D. (Eds). *On the Road*. Almqvist & Wiksell International. pp.176–183.

Price, T. D. et al. (2018). Asnæs Havneemark: a late Mesolithic Ertebølle coastal site in western Sjælland, Denmark. *Danish Journal of Archaeology*, 7 (2), Routledge., pp.255–276. [Online]. Available at: doi:10.1080/21662282.2018.1551979.

Ramsøe, A. et al. (2020). DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *Journal of archaeological science*, 115, p.105080. [Online]. Available at: doi:10.1016/j.jas.2020.105080.

Rowley-Conwy, P. and Zeder, M. (2014). Wild Boar or Domestic Pigs? Response to Evin et al. *World archaeology*, 46 (5), Informa UK Limited., pp.835–840. [Online]. Available at: doi:10.1080/00438243.2014.953712.

Schoch, C. L. et al. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database: the journal of biological databases and curation*, 2020, Oxford University Press (OUP). [Online]. Available at: doi:10.1093/database/baaa062 [Accessed 6 October 2024].

Seiler, M. et al. (2019). Status report of the Trondheim radiocarbon laboratory. *Radiocarbon*, 61 (6), Cambridge University Press (CUP)., pp.1963–1972. [Online]. Available at: doi:10.1017/rdc.2019.115.

Sievers, F. et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7 (1), EMBO., p.539. [Online]. Available at: doi:10.1038/msb.2011.75 [Accessed 4 October 2024].

Sievers, F., Barton, G. J. and Higgins, D. G. (2020). Multiple sequence alignment. In: Baxevanis, A. D., Bader, G. D. and Wishart, D. S. (Eds). *Bioinformatics*. 4th ed. Milton, QLD, Australia : John Wiley & Sons. pp.227–250.

Sievers, F. and Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein science: a publication of the Protein Society*, 27 (1), John Wiley & Sons, Ltd., pp.135–145. [Online]. Available at: doi:10.1002/pro.3290 [Accessed 4 October 2024].

Skousen, H. (2008). *Arkæologi i lange baner: undersøgelser forud for anlæggelsen af motorvejen nord om Århus 1998-2007*. Moesgård.

Sørensen, L. and Karg, S. (2014). The expansion of agrarian societies towards the north – new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. *Journal of archaeological science*, 51, Elsevier BV., pp.98–114. [Online]. Available at: doi:10.1016/j.jas.2012.08.042.

Suzek, B. E. et al. (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics (Oxford, England)*, 31 (6), Oxford University Press (OUP)., pp.926–932. [Online]. Available at: doi:10.1093/bioinformatics/btu739 [Accessed 6 October 2024].

Van Neer, W. et al. (2013). Hunting, gathering, fishing and herding: Animal exploitation in Sandy Flanders (NW Belgium) during the second half of the fifth millennium BC. *Environmental Archaeology*, Routledge. [Online]. Available at: doi:10.1179/1461410313Z.00000000032 [Accessed 29 October 2024].

Warinner, C. et al. (2014). Direct evidence of milk consumption from ancient human dental calculus. *Scientific reports*, 4, nature.com., p.7104. [Online]. Available at: doi:10.1038/srep07104.

Yu, F. et al. (2020). Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Molecular & cellular proteomics: MCP*, 19 (9), Elsevier BV., pp.1575–1585. [Online]. Available at: doi:10.1074/mcp.TIR120.002048.

Groß, D. (2024). Unpublished NISP counts of faunal assemblages of Syltholm II shared by Daniel Groß, 18 September 2024.

6 General discussion

The past decade of palaeoproteomic research has seen rapid expansion in the archaeological materials that could be analysed and questions that could be investigated. As a research community, we have pushed the frontiers of what was deemed possible, such as more species that can be taxonomically identified (Peters et al. 2021; Dierickx et al. 2022), faster (and thus cheaper) analyses (Blacka et al. 2024; Koenig et al. 2024) and new materials that can be analysed (Demarchi et al. 2020; Gianvincenzo et al. 2023; Wilkin et al. 2023). Yet, in the success story of palaeoproteomics, foodcrusts pose a more unyielding frontier. The work in this thesis is one attempt to push back these boundaries. First, a baseline for the taxonomic resolution achievable by proteomics is established through the analysis of a whale bone tool (Chapter 2). Secondly, the impact of the used foods on the formation of foodcrust and how well the recovered proteins reflect the cooked foodstuffs is analysed (Chapter 3). Then, to evaluate the robustness of proteomic taxonomic identifications a proteomic blind test using a set of experimental foodcrust produced with a variety of foodstuffs is performed (Chapter 4). Lastly, the insights gained from these experiments were applied to a case study on Neolithic-Mesolithic foodcrusts from the Danish site Syltholm II (MLF00906-I/II).

The analysis of the whale bone tool (Chapter 2) demonstrated that of the tested workflows only the targeted database search was able to distinguish between closely related species. However, this requires good preservation as well as prior untargeted analysis to establish the taxonomic group for the targeted analysis to focus on.

The first experiment (Chapter 3) set out to investigate how fairly the recovered proteins from a foodcrust reflect the cooked foodstuffs and if the composition of the ingredients influences the amount of foodcrust that forms. The results showed that the protein-rich ingredient, salmon in this case, was indeed overrepresented relative to the other foodstuffs in the mixture (beetroot and beef fat). The observed protein abundance, measured as the sum of intensities, correlated with the total protein content and carbohydrate content. Moreover, combinations of the protein-rich food (salmon) and carbohydrate-rich food (beetroot) also seemed to result in the largest volume of foodcrust. The correlation between protein abundance and the protein content of the cooked ingredients was expected, but the carbohydrate correlation less so. Additionally, the correlation between abundance and protein content does not seem to translate to a correlation between the number of identified proteins and protein content. The only statistically significant correlation observed with the number of identified proteins was a negative correlation with the number of cooking events, i.e. when a vessel is used repeatedly the number of food proteins detected decreases. Additional cooking events and burial also reduce the amount of samples from which all used ingredients could be identified. In fact, after burial, only salmon and beetroot proteins could be recovered. All evidence of the presence of beef had disappeared in the samples obtained after burial. The inability to recover all cooked ingredients and the low number of proteins in modern experimentally generated foodcrust is an ill omen for the analysis of archaeological foodcrust. It simultaneously warns of the challenges as well as sets the expectations that we likely only identify a subset of the foodstuffs that were actually used to form the foodcrust.

It is well known that not all species are equally likely to be identified by proteomics (Pipart et al. 2024). Partially because of different levels of protein content, as shown by the previous experiment, but also because the databases used in MS/MS analysis do not

encompass all species. The second experiment was a blind test (Chapter 4). Twenty foodcrusts, produced during two cooking experiments were analysed and interpreted, blind to the true composition of the samples. The only information known prior to analysis was that the used foodstuffs would likely have been available to Danish Mesolithic hunter gatherers, in other words only constrained by the same assumptions used when interpreting material culture from a site. The aim of this experiment was to establish which species that could be part of archaeological foodcrusts can actually be identified and which species are likely to remain unidentified, despite their presence.

Additionally, we compared two extraction protocols (SP3 and DES), two mass spectrometers (Orbitrap Fusion Tribrid and Q Exactive HF Orbitrap, both Thermo Scientific) and several bioinformatic tools for taxonomic validation (Unipept, ProteoClade and BLAST).

The results of the blind test were humbling, but have important ramifications in the field of palaeoproteomics. Similarly to the first cooking experiment not all foodcrusts yielded proteins and the number of proteins that could be recovered from the other samples was limited. However, of greater concern is that in nine of the twenty samples, across all combinations of extraction method and mass spectrometer, no proteins congruent with any of the cooked foodstuffs, regardless of their taxonomic specificity, were found. Furthermore, in 14 of the 20 samples proteins were assigned to a taxon that was *not* part of the ingredients used to produce the foodcrust. These results were obtained using a variety of validation methods and standard identification criteria. Most of the false positives could be removed by requiring a minimum of six distinct peptide sequences per taxon and by summing all remaining taxonomic identifications to order level. These two (extreme) measures reduce the information obtained from proteomic analysis drastically, but the taxonomic identifications obtained with commonly accepted identification criteria would have missed many of the taxa present in the samples and conjured up several other taxa that were never present.

The previous two experiments offer few reasons for high hopes regarding the archaeological case study from Sylthom II and indeed out of 76 foodcrust samples 11 yielded proteins that withstood the validation process. The identified proteins suggest the cooking of dairy (likely curd products) in both the Mesolithic and Neolithic foodcrusts. Moreover, *Xenopus laevis* and *Pisum sativum* proteins were identified in some of the Neolithic samples, but their authenticity remains somewhat questionable. As poor protein recovery was expected, the same dataset was also analysed with *Scanning Electron Microscopy* (SEM) and lipid analysis. The lipid analysis also found evidence of dairy consumption in the Ertebølle (EBK) and Funnelbeaker (TRB), corroborating the protein evidence. Additionally, substantial evidence for the use of marine resources was found in both cultures, as well as some samples that likely contained mixtures of dairy and marine resources. The SEM analysis also found evidence of *Hordeum vulgare* and *Triticum* sp. in the TRB samples, while samples from both cultures contained evidence of wild tubers. Hence, our analysis found evidence of continuity in diet from the Mesolithic to the Neolithic, including marine resources, wild tubers and dairy. The chapter also hints at the role palaeoproteomics can play in dietary analysis alongside methods like lipid analysis and SEM; a relatively low percentage of samples yields data, but identified proteins can add more detail regarding the cooked tissues. Here again the protein data also contains elements that cannot be easily authenticated, raising questions on how to integrate the results with our current understanding of dietary changes during the Mesolithic-Neolithic transition.

In summary, the prime message of all the foodcrust analyses in this thesis is to be cautious with the interpretation of proteomics data. This is not to say that the proteomic analysis of archaeological foodcrusts or other complex mixtures should be abandoned, but above all else the data presented here suggests that we should take a step back and reconsider the assumptions and requirements underlying our methods. In particular, what is the amount of evidence required to be confident in a protein or taxonomic identification. Does the answer lie in the abundance of PSMs or peptide sequences or rather in strict vetting of the quality of the obtained data? Some of the issues discussed here are specific to foodcrusts, but others are general to palaeoproteomic analysis.

The challenges that will be discussed in this chapter can be divided into two categories: a) the nature of foodcrusts and b) bioinformatic analysis. Unresolved questions regarding both subjects remain the primary barrier to maximising their use for palaeodietary analysis. The current role proteomic analysis plays in reconstructing past diets, particularly in relation to other commonly used methods of analysis, will be the third focus point.

6.1 The nature of foodcrust

Palaeoproteomics has been applied to many different materials, but at its root it has been developed in order to investigate biomineralised tissues (Ostrom et al. 2006; Cappellini, Collins and Gilbert 2014). The historical focus on biomineralised tissues, in particular bone, is no accident. The combination of a sizeable organic fraction mostly composed of proteins encased in a mineral matrix shielding it from degradation makes it a prime target for proteomics. Additionally, the composition of the bone proteome has been studied extensively. We know that collagens make up 70% of bone proteins and thus present a suitable target for peptide mass fingerprinting. We know which non-collagenous proteins are most abundant and that the composition of bone is fairly consistent across elements and taxa. It is our knowledge of the biochemical composition and structure of a material such as bone that enables us to extract and identify proteins from heavily degraded samples stretching back millions of years.

So, what then is foodcrust?

As foodcrust is formed from food, it is composed of the same classes of biomolecules as food, meaning that foodcrusts are primarily a mixture of carbohydrates, proteins and lipids. Additionally, the heating involved in the formation of foodcrust leads to chemical changes of these biomolecules and the creation of degradation products, as discussed in section 6.1.2. It is unclear if there is a general pattern in the relative proportions of these molecular classes. It is assumed that the ratio of carbohydrates, proteins and lipids depends on the abundance of these biomolecules in the foods cooked to produce the foodcrust, but it is not known how the foodcrust formation processes affect these ratios. The experiment in chapter 3 aimed to test how differences in the relative abundance of carbohydrates, proteins and lipids might affect the protein recovery and abundance of foodcrust and the results seem to indicate that the amount of protein and carbohydrate are most important for predicting protein recovery. But the abundance of carbohydrates, proteins and lipids was not quantified in the foodcrust after cooking. It may be that the relative abundance between these changed during the foodcrust formation. It has been proposed that Attenuated Total Reflection - Fourier Transform Infrared (ATR-FTIR) spectroscopy might be able to quantify the levels of carbohydrates, proteins and lipids (Ferro et al. 2019). ATR-FTIR has been applied previously to archaeological foodcrusts, although not yet quantitatively (Ge et al. 2024). Future

applications of this method might improve our understanding of the composition of archaeological foodcrust.

Apart from the uncertainty regarding the biomolecular composition of foodcrust, some inconsistency in the terminology used to describe foodcrusts, as mentioned at the beginning of this thesis. Some of the terms used, such as simply 'organic residues' are also applied to other types of material, like biomolecules entrapped in the ceramic matrix.

6.1.1 A different beast by a similar name

In a similar vein, not all materials described as "food crust" appear to be the same. The use of similar terminology for different objects is equally problematic, as it either prevents the lessons learned from other studies being applied by the wider research community or leads to incorrect assumptions based on misunderstandings regarding the analysed materials. The potential risk of terminological inconsistency is illustrated by the following case studies:

Hemudu is a site in East China dated to 7000-5300 BC that yielded ceramic vessels with grey, brown and yellowish residues on the interior surface (Fig. 1A). The residues were water soluble and elemental composition analysis revealed that carbon, oxygen, phosphorus and calcium made up 83.4-97.1% of the elemental content. SEM analysis showed the presence of spongiform structures, which were interpreted as bone particles, while FTIR analysis showed spectral bands congruent with carbonate hydroxyapatite as well as weak absorption bands associated with collagen. Taken together, the vessels were interpreted as having been used to cook bone powder (Ge et al. 2021).

A similar set of "beige residues" (Fig. 1C) is reported from Neolithic Bulgarian sites (Vieugué et al. 2015). FTIR analysis indicates the presence of carbonyl, phosphate and carbonate groups, while elemental composition analysis suggests that the residues are mainly composed of calcium and phosphorus. The chemical composition of the beige residues was said to be different from taphonomic carbonate concretions present on some shards from the same sites, which lacked phosphorus. The beige residues were therefore also interpreted as the result of the cooking of bone powder, although SEM analysis showed no bone inclusions (Vieugué et al. 2015).

A third study on white "mineralised organic residues" (Fig. 1D) from 3rd-7th century BC Italy also obtained FTIR spectra showing hydroxyapatite and low intensity amide peaks. The residues were analysed with LC-MS/MS as well, which found evidence of collagen- α -1(I) and -2(I), but no other commonly identified bone proteins. Dairy proteins were identified in two samples. The residues were again interpreted as being the result of bone processing (Siano et al. 2022).

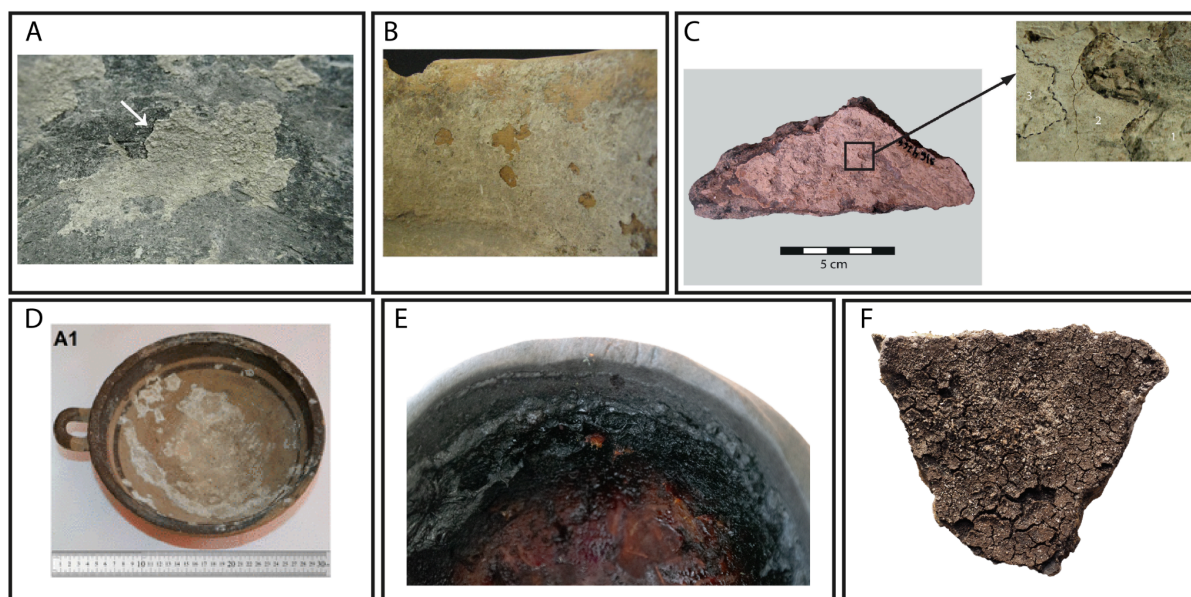


Figure 1, Examples of calcified deposits and “bone powder” residues. A, foodcrust, Fig. 2B(Ge et al. 2021), B, calcified deposit, Fig. 2B (Hendy et al. 2018b), C, beige residue Fig. 4 (Vieugué et al. 2015), D, mineralised organic residue, Fig. 2 (Siano et al. 2022), E, experimentally generated foodcrust, F, EBK foodcrust from Syltholm II.

These studies are great examples of combining multiple methodologies to study organic residues, but a white-grey-brown residue mostly composed of calcium carbonate sounds more similar to “calcified deposits” (Fig. 1B) (Hendy et al. 2018b) rather than “foodcrust” (Heron and Craig 2015), especially when compared to experimental (Fig. 1E) and archaeological foodcrusts (Fig. 1F) that have been analysed in this thesis.

Apart from visual differences, the usage of foodcrust versus calcified deposit also has interpretative implications. The calcium carbonate in calcified deposits is thought to come from the water used for cooking rather than bone powder and this would substantially change the interpretation of the residues. The presence of collagen and hydroxyapatite seems to support the interpretation of the processing of bone, but collagen can also derive from other tissues such as meat and hydroxyapatite has been identified as fouling in modern industrial equipment, including from a sugar factory (East et al. 2011), from milk processing (Rosmaninho et al. 2007) and from a wastewater treatment (Zhang et al. 2021). The bone-like inclusions observed with SEM analysis in the Hemudu residues likely do represent bone, but do not require large amounts of bone powder. The data obtained from the residues in these three studies does not exclude the cooking of bone powder, but neither does it require bone powder. Without controlled experiments on the formation of organic residues as a result from bone powder cooking, calcified deposits formed by the cooking of meat are an equally likely explanation. The correct interpretation of these residues depends greatly on what they exactly are and how they were formed. The most important issue raised here is not the interpretation of the residues, but the fact that the lack of consistency in the description and methods used to characterise foodcrusts and other forms of organic residues across the *entire* subfield makes it difficult to establish which objects belong to the same category and are therefore likely the result of the same formation processes.

6.1.2 Chemical composition

Although further research is required before our understanding of foodcrust matches that of tissues such as bone and enamel, we are not completely in the dark. Foodcrusts are first and foremost organic mixtures. They form out of foodstuffs as the result of heating, often for the purpose of cooking. There are currently no commonly accepted techniques to establish the temperature and duration of the cooking event nor how many cooking events led to the formation of the foodcrust. However, cooking experiments have shown that ω -(*o*-alkylphenyl)alkanoic acids (APAAs), which are frequently observed in the lipid analysis of foodcrusts (Bondetti et al. 2020), only form after heating unsaturated fatty acids at 270°C for one hour or 200°C for five hours (Bondetti et al. 2021). Additionally, the microscopic structure of foodcrusts often also indicates extensive heating (Courel et al. 2020; Heiss et al. 2020), but this does not apply to all foodcrusts as in some case FTIR spectra still show evidence for the presence of well-preserved polysaccharides (Kubiak-Martens et al. 2024). Modern cooking experiments have shown that foodcrusts do not form quickly. Experiments with open fire cooking showed that foodcrusts only started to form after 30-90 minutes (Philippsen 2012), while the cooking experiment in Chapter 2 produced very little foodcrust after one hour of heating at 270°C. Only after several cooking events did substantial levels of foodcrust, similar to what is observed on archaeological material, appear.

As a consequence of the prolonged heating at such high temperatures a number of chemical changes to the cooked foodstuffs can be expected. Proteins are expected to denature and oxidise (Bax et al. 2012), leading to the breakdown of cysteine bridges (Hu et al. 2017). Heating also leads to the loss of carboxyl groups, while basic groups appear relatively unaffected (Yu et al. 2017). Proteins and starch also start to break down to building blocks and smaller aromatic compounds. At 310°C, which is achievable with open fire cooking (Dimoula et al. 2019), nearly all of the starch and protein content is thought to have been converted to stable aromatic compounds (Braadbaart et al. 2004; Nitsch, Charles and Bogaard 2015). 315-323°C is also the temperature range at which significant collagen breakdown was observed with thermogravimetric analysis (Marcilla et al. 2011; Bañón, García and Marcilla 2021).

A substantial part of the proteins and sugars loss is due to Maillard reactions resulting in the formation of a wide range of compounds, including aldehydes, ketones, dicarbonyls, acryl amides, heterocyclic amines, melanoidins and advanced glycation endproducts (AGEs) (Wang, Qian and Yao 2011). A consequence of this process is a correlation between heating time and a decrease in hydrolysable amino acids (Styring et al. 2013). A study on glycine heated to 150°C for 24 hours showed that only 25% of the nitrogen content could be detected in the form of an amino acid (Oldenburg et al. 2021). The loss in amino acid content is not indiscriminate, rather threonine, aspartate and serine were observed to decrease at a faster rate after two hours of heating than the other amino acids (Styring et al. 2013). The lipid component of the foodstuffs does not remain inactive during the heating process. Studies indicate that starch, lipids and protein can form ternary structure complexes during heating (Wang et al. 2020). Starch-lipid-protein complexes have not been studied often proteomically and how their formation impacts the survivability or recoverability of proteins is not well known, but starch-lipid-protein complexes show lower enzymatic efficiency than starch-lipid complexes (Zheng et al. 2018). If enzyme access to the protein chains is hindered, this could mean that our ability to extract and identify the proteins is also impacted. Trapped proteins could therefore be an explanation for the low amount of proteins recovered from foodcrusts in the analyses of this thesis (Chapters 3-5).

The other cooking-induced modifications are also expected to have a major impact on protein recovery. While changes such as the loss of amino acids are irresolvable, other modifications may present surmountable obstacles. Oxidation, loss of carboxyl groups and unspecific protein degradation are issues that can be accounted for in bioinformatic analysis. Furthermore, although the entrapment of proteins in starch-lipid-protein complexes may have prevented the detection of proteins up to this point, they might also aid their preservation by shielding proteins from microbial attack. Deconstructing the complexes might release the trapped proteins and could yield valuable information.

6.1.3 Fair representation

It is evident that a lot of work needs to be done in order to unlock the full potential of foodcrusts. Yet, even if future method development will enable us to fully extract and identify the proteome preserved in the foodcrust, will these proteins accurately represent the cooked foods? Chapter 2 indicates that this most likely will not be the case. After burial, all beef proteins had disappeared and only a few beetroot proteins remained. Salmon proteins, it seemed, were more persistent. This begs the question of what determines if a protein is more likely to survive and it seems there is no easy answer. A study on experimental foodcrust searched for a common denominator in the chemical properties of persevering proteins, revealed that the chemical properties of the persevering proteins differed between different taxa and there was no single peptide property which explained why particular proteins persisted in foodcrust (Evans et al. 2024).

There are some proteins that are found relatively frequently across different sample sets. Beta-lactoglobulin (BLG) is one of the most commonly found dietary proteins, especially in dental calculus (Warinner et al. 2014; Charlton et al. 2019) and in calcified deposits (Hendy et al. 2018b; Evans et al. 2023; Tanasi et al. 2021). Its degradation has been relatively well studied (Ramsøe et al. 2021; Palomo et al. 2023; Fonseca, Freeman and Collins 2024). However, in foodcrust the most common proteins appear to be collagen (particularly alpha-2 (I)) and myosin (Shevchenko et al. 2018; Evans et al. 2024; Lyu et al. 2024). Plant proteins have proven more elusive, although they have been found in ceramic samples before (Chowdhury, Campbell and Buckley 2021) and in both the experimental foodcrusts from Chapters 2 and 3, as well as in the archaeological foodcrust presented in Chapter 4. There are several factors that could contribute to the lack of plant proteins in archaeological foodcrusts. First of all, plant foods generally contain less protein than animal foods, meaning that animal foods are expected to be overrepresented. Secondly, most commonly used protein extraction protocols in palaeoproteomics are derived from protocols designed for bone protein extraction. Bone protein extraction protocols require a demineralisation step, usually in the form of incubating the sample in an acid, to release the proteins from the mineral matrix (Warinner, Korzow Richter and Collins 2022). Plant proteins are not protected by a mineral matrix, but by cell walls, which require a different approach such as mechanical disruption (Alvarez and Naldrett 2016).

Lastly, the number of relevant plant species with sizable reference proteomes is far fewer than mammal species. Among 29 plant species identified across 10 EBK sites (Andersen 2009; Price et al. 2001; Holst 2010; Robinson and Harild 2009; Kubiak-Martens 1999, 2002) the average number of reference proteins in the NCBI database is 120. While for 44 mammal species found at 54 EBK sites (Price et al. 2018; Astrup et al. 2021; Andersen 2009; Price et al. 2001; Richter and Noe-Nygaard 2003; Bratlund 1991; Enghoff 2011; Trolle 2013) the average is 40,054 proteins (Chapter 3 SI 2). The paucity of plant

reference proteomes was also the reason why in Chapter 2 the database search included the proteomes of *Beta vulgaris*, *Chenopodium quinoa* and *Spinacia oleracea*. This greatly increased the plant protein identifications, despite *Beta vulgaris* having a “standard” quality reference proteome on UniProt.

These issues make it clear that we should not expect proteins recovered from foodcrusts to perfectly mirror the foods that comprise the foodcrust. Both in terms of the relative abundance of taxa and the likelihood of identifying particular taxa the data is likely to be skewed in favour of protein-rich tissues and taxa with high quality reference proteomes.

6.2 Bottom up or top down

The vast majority of palaeoproteomic studies are bottom-up, meaning that the extracted proteins are digested into peptides using an enzyme and these peptides are used to identify proteins, and species. Top-down proteomics instead focuses on intact proteins and does not feature a digestion step (Smith and Kelleher 2013). One of the primary applications of top-down proteomics is the identification and localisation of PTMs (Brown et al. 2020). An advantage of a top-down approach over bottom-up in studying PTMs is that top-down analysis provides information on which PTMs co-occur on the same molecule. Whereas one cannot be sure if PTMs discovered on different peptides using a bottom-up approach would co-occur on the same intact protein molecule (Fornelli et al. 2018).

As discussed in section 6.1.2 the cooking processes involved in the production of foodcrust are likely to induce additional modifications. A top-down approach could facilitate the detection of the modified proteins and improve protein recovery. Additionally, the more detailed information regarding the modifications might improve our understanding of how proteins are affected by cooking. Beyond PTMs, where peptides are at times shared between multiple proteins and species, intact proteins are easier to identify and their detection would facilitate more precise taxonomic identifications (Catherman, Skinner and Kelleher 2014).

However, a top-down approach is seldom employed in palaeoproteomic studies, primarily because it is expected that few, if any, intact proteins remain in archaeological material. A digestion-free approach is frequently applied for enamel protein extractions, but as enamel proteins are digested *in vivo* (Bartlett 2013), this is not comparable to conventional top-down analyses. Considering that cooking is also expected to lead to fragmentation of the proteins from the food, it is unclear how successful a top-down approach would be for the analysis of foodcrusts, especially archaeological foodcrusts.

6.3 Bioinformatics

The paucity of reference proteomes is one of the key challenges of the bioinformatic analysis of the extracted proteins. The last few years have been marked by rapid diversification in the development and use of the tools used in palaeoproteomic bioinformatics. This extends beyond standard spectrum-matching protein identification software to include specialist tools for example for deamidation (Mackie et al. 2018; Ramsøe et al. 2020) and taxonomic assignment (Gurdeep Singh et al. 2019; Mooradian et al. 2020). One of the reasons for the recent surge in bioinformatic development is the growing realisation that the previously commonly used workflows designed for modern proteomic research do not match well with the added complications of analysing ancient samples.

Among these complications, there are three key challenges: the percentage of peptide-spectrum matches (PSMs), creation of suitable databases and taxonomic validation.

6.3.1 PSMs

Of the various outputs of protein identification software, PSMs are the most fundamental level. At its core, a PSM is a MS¹ peak with an assigned peptide sequence based on its fragment MS² spectrum. All further processed results, from protein identification to intensity-based abundance measurements, are derived from PSMs. In most samples not all peaks can be identified. There are a variety of explanations for a peak remaining unidentified. For example, it could be that the peak is not in fact a protein fragment. Secondly, the protein the peak belongs to might not be included in the database. Alternatively, the peptide might have too many modifications or the relevant modifications were not included in the search settings. A recent study showed that in modern proteomic experiments the percentage of identified spectra sits slightly above 25%, while in palaeoproteomics it ranged between 0.47% and 12.61% (Chiang, Welker and Collins 2024). These low rates of identifications suggest a general issue in proteomics regarding the ability to identify sample contents, which seems to have reached extreme proportions in palaeoproteomics. Therefore increasing the spectra identification rate seems to be of utmost importance. Especially as additional identified spectra will translate into more identified peptide sequences, more sequence coverage and more proteins, which in the end means that we will be able to be more precise in our taxonomic identifications.

Two of the most important choices that affect the spectrum identification rate are the choice of software and the digestion mode (i.e. specific, semi-specific and non-specific). There are different programs available, both free-to-use and commercial, for the identification of proteins. The two most commonly used are Mascot (Perkins et al. 1999) and MaxQuant (Tyanova, Temu and Cox 2016). Both have been in use since the start of palaeoproteomics and are well-regarded. However, a recent comparison between different protein identification software suggested that Mascot and MaxQuant identify a relatively minor part of the spectra compared to other tools, such as FragPipe and pFind (Palomo et al. 2023). The differences were sizeable and the pattern was the same irrespective of the digestion mode. At a FDR (false discovery rate) of 0.01, Fragpipe and pFind identified 2-3 times as many spectra as Mascot and MaxQuant (Palomo et al. 2023).

A shared characteristic of Fragpipe and pFind is that these two programs are capable of performing “open” searches. In classical “narrow” or “closed” searches the user selects a limited number of PTMs (post-translational modifications) that the program will take into account in its protein identification. In “open” searches the user does not select PTMs, but the program will use a much larger search window in order to identify any potential modification. In the case of pFind, the identified PTMs are then immediately used by the program in a following narrow search. The aim of this strategy is to be unbiased towards any PTMs, while maintaining the confidence of the stricter mass tolerance used in a closed search. The same strategy can be applied to Fragpipe, but it requires manually selecting the PTMs found by the open search.

Open search software often seems to outperform traditional closed searches (Kong et al. 2017) and its logic is appealing, particularly for archaeological samples and even more so for complex mixtures. For example the deamidation of N and Q is always selected as a PTM and is used as a marker of protein preservational quality. However, the age-related

diagenesis is likely to produce additional not-well characterised modifications and non-specific peptide cleavage. Cooking, as discussed in the previous section and Chapter 1, is also known to induce many modifications. These modifications would be too many to include in a closed search, but in theory an open search would take all these unknown modifications into account.

An unbiased approach to PTM detection is expected to increase the fraction of identified spectra, but it will not help to identify peptide sequences from proteins that are not included in the used protein database. However, another innovation in protein identification software might: *de novo* sequencing. *De novo* software does not *initially* depend on a database. Instead it calculates the most likely amino acid sequence of a peptide just based on its and its fragments' mass. The appeal of this approach to palaeoproteomics is that it could help in identifying species without good quality reference proteomes. Furthermore, the incorporation of transformer models in recently developed software has led to a significant improvement in the performance of *de novo* software (Yilmaz et al. 2022, 2024).

However, *de novo* sequencing has its own limitations. A key issue is that *de novo* software has the tendency to 'hallucinate' by observing peptide sequences that do not exist (Muth and Renard 2018). Some of these are easily identifiable by the user, for example peptides only consisting of repeats of the same amino acid. To improve the reliability of *de novo* sequences some software tools perform a follow-up closed search. Although this combines the confidence of closed searches with less restricted spectra matching of *de novo* it does tie the *de novo* sequences to a database and loses its database independence as a consequence.

Each of the discussed types of protein searches has its own advantages and disadvantages. In Chapters 4 and 5 the used software, which included *de novo*, open and closed searches, was chosen to maximise the number of PSMs. The samples analysed in both chapters also included proteins that were not present in the used databases. In Chapter 4 many false positives of proteins belonging to taxa that were not present in the samples were identified. The archaeological nature of the samples of Chapter 5 makes it more difficult to ascertain if a protein identification is a true or false positive, but the results did contain some protein identifications of questionable nature. As discussed in Chapter 4, it may be that the software's determination to assign an identity to as many spectra as possible combined with the used database lacking an entry of the protein's "true identity" results in spectra being shoehorned into incorrect protein and species identifications. The likelihood of these errors occurring is unknown, but it may explain some of the misidentifications observed in the blind test in Chapter 4.

6.3.2 Databases

The issue of incomplete databases merits further discussion. There are a number of curated publicly available protein sequence databases. The two most popular of these are UniProt (UniProt Consortium 2023) and NCBI (Sayers et al. 2022). Of the two, NCBI is more extensive. BLAST (Altschul et al. 1990), which is often considered the industry standard for taxonomic validation, uses the NCBI non-redundant database for its searches. Nevertheless, there are blind spots in these publicly available databases, including NCBI. The proteome coverage of domesticated species, model organisms and a number of other species that have been the focus of conservation research are of high quality, but for many wild species the reference sequences are sparse. One must remain mindful of the fact that the primary use of these resources is not archaeology, but medical or commercial research and that is

reflected in which species have high quality reference proteomes. Unfortunately the issue of incomplete references is not something that can be solved easily or quickly. In order to obtain high quality reference proteomes, the species' genomes must first be sequenced, which can then be translated into a proteome. Ideally, the translated proteome is then verified using *de novo* mass spectrometry on controlled samples. There are currently several projects ongoing that aim to map the present biodiversity by sequencing the genomes of large numbers of species (Darwin Tree of Life; Earth Biogenome Project 2022; ERGA). As genomes start to become available from these projects it will be up to the palaeoproteomics community to ensure their translation and validation into proteomes.

Considering the already limited coverage of species and proteins in the current available reference databases, it might seem that the best course of action would be to search the obtained mass spectra against an as large as possible database spanning as many taxa as possible. However, this approach has been found to actually reduce the number of PSMs (Li et al. 2016). This reduction in PSMs is an accidental consequence of a quality control feature called FDR. The FDR is calculated by first creating a "decoy" database. The decoy database is a copy of the used protein sequence database, but all protein sequences are reversed or randomised. This creates a database of fake sequences that remain chemically sound. The real and decoy databases are mixed for the protein search during which each PSM is scored. A threshold FDR value is set, usually 0.05 or 0.01, which represents the percentage of PSMs that are allowed to be decoy matches. Based on the FDR threshold a score cutoff is set and all PSMs below the score cutoff value are discarded. The theory behind this control is that any PSM with a score lower than the cutoff value is not reliable enough to use in downstream analysis. The score value is based upon a number of factors such as fragment ion coverage and most protein identification software use their own scoring algorithm.

A principal assumption of the FDR model is that every peptide is equally likely to match with a PSM, yet this assumption is frequently violated in reality. If a database contains many entries of proteins that are not actually present in the sample, the chance that a spectrum is matched to a decoy increases relatively to the chance that a sample peptide matches to its entry in the database. Consequently, the number of decoy matches increases and inflates the score threshold causing more "true" matches to be discarded (Jeong, Kim and Bandeira 2012; Noble 2015). Especially for samples such as foodcrust, which do not contain much protein to start with, this can have a significant impact on the amount of information that can be obtained from a sample. A commonly used database in palaeoproteomics for complex mixtures without any prior knowledge of their contents is SwissProt, which is a curated database spanning the tree of life (Boutet et al. 2007). Due to SwissProt's size and the taxonomic diversity it contains, the database is guaranteed to contain many entries from species that are not present in the samples. Consequently, using SwissProt may reduce the number of PSMs with an acceptable score.

The alternative could be to use small constrained databases of a number of selected proteins and taxa. This is frequently done in studies where one knows the species of the sample and what proteins to expect, such as in the cooking experiment in Chapter 2. However, when analysing archaeological foodcrust one cannot make any prior assumptions about the contents of the sample. Using other methods such as SEM or lipid analysis, as done in Chapter 4 and see e.g. Lyu et al. (2024), is in general advised, but cannot be used to argue for the exclusion of taxa from a database. Doing so would heavily increase the risk of a confirmation bias, especially considering the aforementioned risk of a spectrum being shoehorned into an incorrect peptide identification if its true identity is not an option.

In the end it seems that the researcher is stuck between a rock and a hard place. Either the used database is too wide, leading to reduced PSMs and a loss of information, or the database is too narrow and one might miss parts of the samples' contents.

6.3.3 Taxonomic validation

In the research presented in this thesis, regardless of the database used, I have found that the initial taxonomic identification proposed by many different protein search engines is nearly always incorrect. The blind test of Chapter 2 illustrates this perfectly. The difficulty these programs have with correctly identifying the taxon partially lies in how the scoring algorithm works. In the end every spectrum is assigned to one peptide belonging to one protein even though the peptide itself can be shared among multiple proteins and taxa. Some software, such as Mascot and MaxQuant, use protein groups to group together peptides shared between similar proteins, but these often do not fully capture the taxonomic width at which a peptide is shared.

Instead there are other programmes specialised in determining the specificity of a given peptide sequence, such as Unipept (Gurdeep Singh et al. 2019) and ProteoClade (Mooradian et al. 2020). Each of the two has its own advantages. Unipept is incredibly user-friendly and due to its advanced missed cleavage handling it assigns a taxonomic level to nearly all peptide sequences. Unipept's main disadvantage is that it can only use UniProt or UniProt-derived databases, which often lack the taxonomic specificity desired in palaeoproteomic research. ProteoClade on the other hand allows the use of any custom database. It allows the user to define their own cleavage rules as well providing a great deal of flexibility. The ability to use any database should allow ProteoClade to become equivalent to BLAST, yet Chapter 3 showed that the two programs at times still yielded different results.

The reason for the different taxonomic identifications between BLAST and ProteoClade lies in the manner of how their database is constructed. BLAST uses the online complete NCBI repository, whereas ProteoClade needs to be supplied with a database. Although it is possible to build a database from NCBI, issues regarding gene name annotation, such as not all homologous proteins being annotated with the same gene name, make it difficult to create a database with no missing taxa or proteins as discussed in more detail in Chapter 3. As it stands, BLAST remains the standard, but new tools are being developed such as ClassiCOL (Engels et al. 2024). ClassiCOL, rather than only using the most specific peptides to establish a protein's taxon, uses a decision-tree approach. For example, if a protein has six associated peptides, of which two are specific to Bovidae and two others are shared between *Bos sp.* and *Equus sp.*, it uses the prior information from the Bovidae peptides to assign *Bos sp.* as final taxon. Such an approach seems incredibly promising for maximising the taxonomic information that can be obtained from a sample, but it remains to be seen how well it works in a mixture of different species.

ClassiCOL's alternative approach to taxonomic identification raises the question if the current approach of focusing on a sequence's level of specificity used in Chapters 3 and 4 might not be overly restrictive. By requiring each protein in a sample to have two peptides specific to order level or a more specific taxonomic rank a lot of information was lost. The alpha-s1-casein peptides found in Chapter 4 show that peptides originally validated at the order level of Pecora could be refined to species level identification using the archaeological context. The two peptide requirement is an important criterion to set a common threshold for identification (Hendy et al. 2018a), but it is at its root an arbitrary number intended to remove one-hit wonders. Different ways of counting could be explored to retain more information.

For example, one could only apply the two peptide limit before taxonomic validation, or one could count taxonomic identifications cumulatively, e.g. one *Bos taurus* and one *Bos* sp. peptide would result in a valid *Bos* sp. identification. One could group the counting per taxon and allow two peptides from different proteins to add up to a valid taxonomic identification. Alternatively, one could repeat the taxonomic validation in a similar way to ClassiCOL. For example, if BLAST were to indicate that of the six peptides of a protein three were specific to Cervidae, one could rerun the six peptides on BLAST but limiting the database to Cervidae only and repeat until the taxon can no longer be refined.

Undoubtedly, if the data from Chapter 4 were to be reanalysed according to any of the above strategies a number of additional samples, proteins and taxa would meet the identification criteria and refine our interpretation of dietary changes and continuity during the Mesolithic-Neolithic transition. However, the identification criteria maintained in Chapter 3 were even more strict than in Chapter 4. And if the criteria from Chapter 3 were to be applied to Chapter 4, only sample X6340 would still yield protein identifications. Considering the reason for the strict criteria in Chapter 3 was the need to remove false positives, any enthusiasm for relaxing protein identification criteria should perhaps be tempered. The reason for relaxing the minimum of six distinct peptide sequences, as suggested in Chapter 3, in Chapter 4 was that due to their age the archaeological samples would be more degraded. Lower peptide counts were expected both for any true protein matches and any false positives. The identified proteins match with the results from the lipid analysis, so it could be that these criteria were sufficient for archaeological samples, but additional research is required to gain more confidence in proper identification criteria.

6.3.4 A solution? Sequential searching

There are plenty of issues and challenges in the palaeoproteomic analysis of complex mixtures. For many of these challenges no clear solution seems in sight or, as in the case of reference genomes, we are dependent on other projects focused on generating reference sequences/proteomes. Yet, some improvements might be made to the workflow. Perhaps the different strengths of *de novo*, open and closed searches can be harnessed by applying all three to the same dataset sequentially. Such an approach would combine the maximised PSM identification from *de novo*, the PTM discovery from open searches and the confidence from closed searches. This strategy could be implemented by combining any three programs, but a combined pipeline called Cerberus is currently being developed (Chiang 2024). The future will show how well it performs.

Additionally, the above process could be repeated twice starting with a wide database and followed by a more restricted database, similar to Evans et al. (Evans et al. 2023). The second database would include homologous proteins for the relevant taxa from all identified proteins. By applying a wide database first, this will hopefully provide a suitable overview of all proteins present in the sample, while the more restricted database increases the number of peptides identified. Some proteins may be missed in the initial database search due to the FDR score inflation, but these will have had the lowest scores and thus been the least reliable in any case. Double searching everything will increase the work required per sample, but considering the small amounts of protein identified in foodcrust so far, it may be worthwhile.

A similar wide and narrow approach could be applied to taxonomic validation. The first run of validation would be used to establish which taxa are present in the samples, the second how much support there is for each taxon across a number of samples. The second

search would also provide the opportunity to remove taxa from the database used for validation that are known to not have been present at the time of the sample's formation. Such a filter step would help reduce the bias towards model organisms that are prone to skewing the results and result in more specific taxonomic identifications.

6.4 The place of proteomics in the archaeology of past diet

Many challenges for the analysis of proteins of foodcrust have been detailed in the previous sections and throughout the thesis. Yet in the application of new methods challenges are to be expected. A more important question is whether proteomics lives up to its promises. Three reasons that are often given for the use of protein methods to analyse complex mixtures such as foodcrusts are that proteomics is tissue specific, has a high taxonomic resolution and can distinguish between components of a mixture.

6.4.1 Tissue specificity

Proteins can be, to a certain extent, tissue specific. The number of tissues a protein is expressed in depends on their function and differs considerably between individual proteins. Caseins and lactoglobulins are milk-specific for example, while others like ubiquitin are expressed across the body (UniProt Consortium 2023). Tissue specificity annotations must be interpreted with some level of caution, as most proteins are not strictly specific but often simply more intensely expressed in certain tissues (Uhlén et al. 2019), and expression data are often based on mRNA expression rather than direct protein identification. Additionally, the tissue categories listed in tissue atlases are classified for medical rather than dietary or archaeological purposes. For example, the Human Protein Atlas has a category for bone marrow, but not for bone itself, dentine or enamel. This misalignment of interests can complicate correctly interpreting the level of tissue specificity of proteins identified in materials like archaeological foodcrust.

The tissue specificity of proteins may be best evaluated by comparing it against that of lipid analysis, the most common method of analysing foodcrust. In terms of tissues, lipid analysis is able to distinguish between adipose and milk fat. Plant lipids can also be identified, but no distinction can be made between different tissues of the plant. Proteins are expected to be able to distinguish between tissues such as meat, organs, eggs, dairy, seeds and leaves/stalks, but most of the proteins identified in Chapters 2 and 3 are fairly unspecific. Fig. 6 in Chapter 2 shows the tissue association of all proteins identified in the cooking experiment. The vast majority of tissue annotations for both the *Bos taurus* and *Salmo salar* proteins were unspecific, but each also features several organ specific proteins, which were not included in the cooked ingredients.

In Chapter 3 many of the samples were assigned as meat due to the presence of proteins such as myosin, collagen alpha-2(I) and collagen alpha-1(I). The identification of plant greens in BT86 was based on the identification of chloroplastic ferritin, which is present in chloroplasts, but this turned out to be a false positive. Although a plant was cooked in BT86, it was crab apple (*Malus sylvestris*), rather than greens. The samples analysed in Chapter 3 also contained other tissue types such as roe and bird eggs, but no corresponding proteins were identified. It cannot be said with certainty whether the lack of fish roe or bird egg proteins indicates that evidence of these tissues is more difficult to find or if their absence is a mere coincidence in severely degraded samples.

One of the blind test samples, BT74, contained milk, although no dairy proteins were detected. Dairy products are frequently identified in archaeological materials such as dental calculus (Jeong et al. 2018; Wilkin et al. 2021; Bleasdale et al. 2021; Tang et al. 2023) and calcified deposits (Evans et al. 2023), so its absence was worrying. However, the archaeological foodcrust in Chapter 4 did yield dairy proteins, indicating that dairy proteins can also survive in foodcrusts.

In the end the datasets included in this thesis do not allow to resolve the question whether proteomic data extracted from foodcrusts enable tissue specific identifications. In large datasets such as in Chapter 2, incorrect tissue annotations are observed, while in Chapter 3 only proteins associated with the muscle-based foodstuffs were identified. It is not possible to determine if the other tissues would have been correctly identified if proteins from the respective taxon had been recovered. Consequently, from the data in this thesis it is also not possible to determine whether proteins are able to distinguish between more tissues than lipid analysis.

6.4.2 Taxonomic identifications

The second main advantage of proteomics compared to other commonly used methods of dietary identification that is often cited is its taxonomic resolution. One of the first applications of palaeoproteomics was as a way to distinguish between sheep and goat using ZooMS (Buckley et al. 2009) and protein-based taxonomic identification has been refined to even be able to distinguish between members of a genus using bone and enamel samples (Chen et al. 2019; Welker et al. 2020; Paladugu et al. 2023). However, taxonomic identification of samples such as osseous objects and skins has some advantages over mixtures like foodcrust, the most important of which is that they only consist of a single species. Consequently proteins with different levels of taxonomic specificity can be combined to determine the most probable origin (Chapter 2). In foodcrust on the other hand, each protein must be considered on its own, as any number of species could have made up the foodcrust. It is therefore expected that the taxonomic resolution of protein identification will be somewhat reduced in foodcrust, but to what extent? Again, the standard to compare against is set by lipid analysis, which is able to distinguish between marine, freshwater resources, porcine fats, ruminant and non-ruminant adipose tissue, C3 and C4 plants.

Ironically, the proteins identified in the archaeological samples in Chapter 4 were generally assigned to a more specific taxon than the proteins from the experimental foodcrusts in Chapter 3. For the identified dairy proteins it was even possible to obtain a species-level identification. Of course, this is partially due to the reordering of all protein identifications to order level, but this was a necessary adjustment. If proteomics were only able to make order level identifications, its specificity would be very similar to lipid analysis. Key differences would be that pig/boar (*Sus* sp.) would fall in the order of Artiodactyla (even-toed ungulates) and that marine and freshwater fish would no longer be distinguishable, but marine mammals and marine fish would be distinguishable. Proteomic results would be as unspecific as lipid analysis, but the categories would be somewhat different.

However, such drastic filtering might not always be necessary. The identifications presented in Chapter 4 were initially made to various taxonomic levels, but for the dairy peptides could be refined to species level by comparison against a restricted set of possible species. In contrast the *Pisum sativum* and *Xenopus laevis* identifications were marked as species-specific by BLAST but should be interpreted as more generic based on the

archaeological context as well. Regarding the addressed issues concerning taxonomic validation, one may conclude that the palaeoproteomic analysis of foodcrust has the potential to produce more specific taxonomic identifications than lipid analysis, but that this is not guaranteed.

More important than resolution is the accuracy of taxonomic identifications. Archaeology in general, but dietary analysis in particular, is by nature an interdisciplinary endeavour. The combination of data from a variety of methods such as archaeobotany, zooarchaeology, lipid and isotope analysis and proteomics can lead to emergent insights that transcend the limitations of each individual methodology. However, the other side of the coin of interdisciplinarity is that one must expect that some nuance will be lost when experts from different disciplines build on your research to further progress the field. Any taxonomic identifications presented must therefore be unambiguous and reliable.

Regrettably (but importantly), rather than removing ambiguity this thesis might actually sow some seeds of doubt. The blind test in Chapter 3 clearly shows the challenges involved in accurate taxonomic identifications and ironically the paucity of data might improve the reliability of the results from the archaeological data in Chapter 4. The results presented here offer no clear answer on how to improve the accuracy of our taxonomic identification, but perhaps consolation can be found in the fact that 17 years ago 11 labs participated in a blind test to identify the foodstuff that was cooked in a ceramic vessel using lipid, stable isotope analysis and proteomics and none succeeded (Barnard et al. 2007). Despite, or perhaps partially due to this negative result these methods have all been further developed and are routinely applied today. Proteomics may simply need more time and research.

What then is the place of proteomics among other methods for the analysis of archaeological foodcrusts? In terms of tissue specificity and taxonomic resolution it can outperform lipid analysis, but Chapter 5 highlights that this is not a reliable occurrence. However, lipid analysis often requires smaller sample sizes and is cheaper. Similarly, SEM analysis is cheap to perform and can achieve highly precise taxonomic identifications for a variety of plant taxa. Meanwhile, the accuracy of proteomics is not without doubt and protein extractions on archaeological material were only successful in 11 out of 76 samples in Chapter 4. Is it still worthwhile to attempt palaeoproteomic analysis of foodcrust?

On its own, the proteomic analysis of archaeological foodcrusts seems a risky endeavour. The chances of recovering any proteins are slim, the validation process is complex and ambiguous. The final results may lead to new discoveries or simply confirm what other more economical methods had already established. To maximise the chances of success, it seems prudent to first apply other methods of analysis to establish the presence of proteins. For instance, amino acid analysis, bulk isotopic analysis and ATR-FTIR could all provide some supporting evidence for the presence of proteins. Such an approach might increase the chances of successfully retrieving proteins.

Additionally, perhaps data from SEM, lipid analysis or proteomics should not be compared to each other as being more or less specific, but rather as providing information on different aspects of the diet. SEM far outperforms either lipid analysis or proteomics at identifying plant resources, as shown in Chapter 4. Lipid analysis is robust and identifies what contributed the most fatty acids to the foodcrust, while proteomics of course focuses on the proteins. These resources need not be the same and by combining the different techniques a more comprehensive understanding of diet can be obtained.

6.4.3 Reevaluating relevance

At the start of this thesis in section 1.3 three major archaeological debates were identified as fields of research that the proteomic analysis of foodcrust could have a major impact on. These three were:

1. The use of the earliest pottery and consequently the reasons for adopting pottery.
2. Culinary specialisation.
3. Continuity and change in diet and cuisine during periods of cultural changes.

The potential species-specific taxonomic resolution, the ability to distinguish between components of a mixture, as well as identifying the cooked tissue that palaeoproteomic analysis was expected to offer, were expected to enable substantial progress in each of these debates. As discussed in sections 6.4.1 and 6.4.2, the results presented in this thesis suggest that currently palaeoproteomic analysis of foodcrusts falls short of its promises regarding tissue and taxonomic specificity. Consequently, it is necessary to revise what a more realistic contribution of the proteomic foodcrust analysis to these debates could be.

Regarding the earliest use of pottery it remains to be seen if any proteomic data can be obtained. In the archaeological case study presented in chapter 5, from 73 foodcrusts dating to 4000 - 3700 cal. BC, proteins could only be successfully identified in 11 samples. The earliest foodcrusts that have been successfully analysed with lipid analysis date to 13000 - 9200 cal. BC (Craig et al. 2013), making them more than twice as old as the samples analysed in chapter 5. The relationship between protein survival and age is not linear, but considering the low frequency of samples yielding proteins from sites of around 6000 years old, expectations regarding protein survival in samples twice that age must be tempered. Furthermore, even if proteins can be successfully recovered, chapter 3 has shown that often not all components of a food mixture can be identified. Any pattern in pottery use identified through palaeoproteomics is therefore likely biased towards protein-rich resources. However, chapter 3 also shows that lipid-rich foodstuffs are not preferentially represented in the recovered proteins. By combining lipid and protein analysis a more complete view of the use of the earliest pottery might emerge.

The need for a combined approach of multiple methodologies applies to the study of culinary specialisation as well. If foodcrusts are *only* studied with palaeoproteomics, the risk of missing ingredients, as demonstrated by chapter 3, or misidentifying the cooked foods, a significant risk as shown by chapter 4, is too great to confidently discuss the presence of specialisation. However, by combining SEM, lipid and protein analysis, as done in chapter 5, biases towards any type of foodstuff are minimised. With a more holistic understanding of what foods were processed in the ceramic vessels, the question of specialisation can be addressed more effectively.

As for the last debate regarding continuity and change in diet during cultural transitions, chapter 5 aimed to directly contribute to this debate. In particular, chapter 5 focused on the Mesolithic - Neolithic transition in Denmark. As discussed previously, only a limited number of proteins could be recovered and only from a small subset of the analysed samples. However, the proteins that were identified did provide important information regarding a continual use of dairy throughout the studied period. Furthermore, the combination of SEM and lipid analysis showed continuity in the use of aquatic resources and wild plants as well. The only clear change between the EBK and TRB samples was the evidence for cereals, particularly *Hordeum vulgare*, observed through SEM analysis in TRB foodcrusts. The detection of dairy in the EBK adds direct evidence to previous suggestions of the presence of dairy in EBK vessels from the $\delta^{13}\text{C}$ values of stearic and palmitic acid

(Papakosta, Oras and Isaksson 2019; Courel et al. 2020; Lucquin et al. 2023)(Sørensen and Karg 2014)the analyses presented in this thesis have successfully contributed new knowledge on the debate of dietary changes during the Mesolithic - Neolithic transition.

6.5 Future directions

It is clear that much further work is needed in order to understand what foodcrusts are and how they can best be used in archaeological analysis. A number of specific suggestions have already been made in previous sections and the discussions of the other chapters, but perhaps most importantly a change of strategy is required. At the start of section 6.1 on *The nature of foodcrust* I drew a comparison between bone and foodcrusts and what made bone so suitable for palaeoproteomic research, while foodcrust has proven a much more challenging material. The core difference between the two is the level of our understanding of the biochemical composition of the material. Fundamental questions regarding foodcrust, such as what the requirements are for its formation, what molecules it is composed of, how these are bound and how those bonds can be broken, remain unanswered. Without that basic knowledge, attempts at method development, such as the adapted DES protocol used in Chapters 3 and 4, will be inefficient.

To enable meaningful progress to be made, the foundations of our understanding of foodcrusts must first be secured. Proteomics should not be the first step in the process. The challenges outlined above add too much complexity to the interpretation. Instead, methods such as amino acid composition analysis, elemental composition analysis and FTIR should be among the first methods applied to foodcrust. Amino acid composition analysis will show whether amino acids are present at all and if their relative ratios are resemblant to bacterial contamination. Additionally, if at a later stage no proteins can be recovered from the same samples, this will indicate that proteinaceous material was present but has simply been degraded too much. Elemental composition analysis and FTIR will provide useful information on the chemical composition of the samples and can be easily and non-destructively applied to archaeological samples, facilitating the comparison between archaeological and experimental samples. It is only after these fundamental experiments have been performed that proteomic analysis should be considered.

More complex chemical experiments will also be required to investigate the formation of lipid-carbohydrate-protein complexes and how these might either benefit or hinder proteomic analysis. Breaking up these complexes might release proteins for proteomic analysis. Furthermore, it is worthwhile investigating if these complexes along with Maillard reaction products could be used to investigate the cooking process.

Considering all the work required in order to improve the extraction and interpretation of proteins from foodcrusts, other materials might be considered as alternative targets of analysis. For example, calcified deposits seem more compliant in yielding proteins (Hendy et al. 2018b; Evans et al. 2023). However, many of the questions that complicate the interpretation of proteins from foodcrust also apply to calcified deposits, such as the accuracy of taxonomic assignments from a complex mixture of proteins.

Lastly, the results of the Chapter 3's blind test will remain a thorn in the side of the interpretations of palaeoproteomic identifications until either a new blind test is performed or the data from Chapter 3 is reanalysed using new tools or programs. Development in bioinformatic workflows may contribute substantially to resolving the challenges of analysing complex mixtures. In fact, I hypothesise that the reanalysis of all published palaeoproteomic

data, in particular from studies on complex mixtures, with software developed before the end of the decade is likely to significantly change our interpretation of the data and might either refine our current taxonomic identifications or reveal evidence of the exploitation of new resources. The aim of this thesis was to push back the boundaries of the palaeoproteomic analysis of foodcrust, but perhaps its main contribution to the field lies rather in mapping the borders of what is currently possible and highlighting both challenging obstacles and promising avenues of further research.

References

- Altschul, S. F. et al. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215 (3), pp.403–410. [Online]. Available at: doi:10.1016/S0022-2836(05)80360-2.
- Alvarez, S. and Naldrett, M. J. (2016). Plant structure and specificity - challenges and sample preparation considerations for proteomics. *Advances in experimental medicine and biology*, 919, Springer International Publishing., pp.63–81. [Online]. Available at: doi:10.1007/978-3-319-41448-5_4 [Accessed 30 October 2024].
- Andersen, S. H. (2009). *Ronæs Skov, Marinarkæologiske undersøgelser af en kystboplads fra Ertebølletid*. Højbjerg : Jysk Arkæologisk Selskab.
- Astrup, P. M. et al. (2021). A drowned Mesolithic shell midden complex at Hjarnø Vesterhoved, Denmark and its wider significance. *Quaternary science reviews*, 258, p.106854. [Online]. Available at: doi:10.1016/j.quascirev.2021.106854.
- Bañón, E., García, A. N. and Marcilla, A. (2021). Thermogravimetric analysis and Py-GC/MS for discrimination of leather from different animal species and tanning processes. *Journal of analytical and applied pyrolysis*, 159, p.105244. [Online]. Available at: doi:10.1016/j.jaap.2021.105244.
- Barnard, H. et al. (2007). Mixed results of seven methods for organic residue analysis applied to one vessel with the residue of a known foodstuff. *Journal of archaeological science*, 34 (1), pp.28–37. [Online]. Available at: doi:10.1016/j.jas.2006.03.010.
- Bartlett, J. D. (2013). Dental enamel development: proteinases and their enamel matrix substrates. *ISRN dentistry*, 2013 (1), Hindawi Limited., p.684607. [Online]. Available at: doi:10.1155/2013/684607 [Accessed 21 October 2024].
- Bax, M.-L. et al. (2012). Cooking temperature is a key determinant of in vitro meat protein digestion rate: investigation of underlying mechanisms. *Journal of agricultural and food chemistry*, 60 (10), pp.2569–2576. [Online]. Available at: doi:10.1021/jf205280y.
- Blacka, C. et al. (2024). Rapid amelogenin sex estimation of human and cattle remains using untargeted Evosep-timsTOF mass spectrometry. *Authorea Inc.* [Online]. Available at: doi:10.22541/au.172599306.61984507/v1.
- Bleasdale, M. et al. (2021). Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nature communications*, 12 (1), nature.com., p.632. [Online]. Available at: doi:10.1038/s41467-020-20682-3.
- Bondetti, M. et al. (2020). Fruits, fish and the introduction of pottery in the Eastern European plain: Lipid residue analysis of ceramic vessels from Zamostje 2. *Quaternary international: the journal of the International Union for Quaternary Research*, 541, pp.104–114. [Online]. Available at: doi:10.1016/j.quaint.2019.05.008.
- Bondetti, M. et al. (2021). Investigating the formation and diagnostic value of ω -(o-alkylphenyl)alkanoic acids in ancient pottery. *Archaeometry*, 63 (3), Wiley., pp.594–608. [Online]. Available at: doi:10.1111/arcm.12631.
- Boutet, E. et al. (2007). UniProtKB/Swiss-Prot. *Methods in molecular biology (Clifton, N.J.)*, 406, Methods Mol Biol., pp.89–112. [Online]. Available at: doi:10.1007/978-1-59745-535-0_4 [Accessed 14 October 2024].
- Braadbaart, F. et al. (2004). Laboratory simulations of the transformation of peas as a result

of heating: the change of the molecular composition by DTMS. *Journal of analytical and applied pyrolysis*, 71 (2), Elsevier BV., pp.997–1026. [Online]. Available at: doi:10.1016/j.jaap.2004.01.001 [Accessed 30 October 2024].

Bratlund, B. (1991). The bone remains of mammals and birds from the bjørnsholm shell-mound. *Journal of Danish archaeology*, 10 (1), Det Kgl. Bibliotek/Royal Danish Library., pp.97–104. [Online]. Available at: doi:10.1080/0108464x.1991.10590055.

Brown, K. A. et al. (2020). Top-down proteomics: challenges, innovations, and applications in basic and clinical research. *Expert review of proteomics*, 17 (10), Informa UK Limited., pp.719–733. [Online]. Available at: doi:10.1080/14789450.2020.1855982.

Buckley, M. et al. (2009). Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 23 (23), Wiley., pp.3843–3854. [Online]. Available at: doi:10.1002/rcm.4316.

Cappellini, E., Collins, M. J. and Gilbert, M. T. P. (2014). Biochemistry. Unlocking ancient protein palimpsests. *Science (New York, N.Y.)*, 343 (6177), American Association for the Advancement of Science (AAAS)., pp.1320–1322. [Online]. Available at: doi:10.1126/science.1249274.

Catherman, A. D., Skinner, O. S. and Kelleher, N. L. (2014). Top Down proteomics: facts and perspectives. *Biochemical and biophysical research communications*, 445 (4), Elsevier BV., pp.683–693. [Online]. Available at: doi:10.1016/j.bbrc.2014.02.041.

Charlton, S. et al. (2019). New insights into Neolithic milk consumption through proteomic analysis of dental calculus. *Archaeological and anthropological sciences*, 11 (11), Springer., pp.6183–6196. [Online]. Available at: doi:10.1007/s12520-019-00911-7.

Chen, F. et al. (2019). A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature*, 569 (7756), nature.com., pp.409–412. [Online]. Available at: doi:10.1038/s41586-019-1139-x.

Chiang, Y., Welker, F. and Collins, M. J. (2024). Spectra without stories: reporting 94% dark and unidentified ancient proteomes. *Open research Europe*, 4, p.71. [Online]. Available at: doi:10.12688/openreseurope.17225.1.

Chowdhury, M., Campbell, S. and Buckley, M. (2021). Proteomic analysis of archaeological ceramics from Tell Khaiber, southern Iraq. *Journal of archaeological science*, 132, p.105414. [Online]. Available at: doi:10.1016/j.jas.2021.105414.

Courel, B. et al. (2020). Organic residue analysis shows sub-regional patterns in the use of pottery by Northern European hunter–gatherers. *Royal Society Open Science*, 7 (4), Royal Society., p.192016. [Online]. Available at: doi:10.1098/rsos.192016.

Craig, O. E. et al. (2013). Earliest evidence for the use of pottery. *Nature*, Nature Publishing Group. [Online]. Available at: doi:10.1038/nature12109.

Darwin Tree of Life. *Darwin Tree of Life*. [Online]. Available at: <https://www.darwintreeoflife.org/> [Accessed 19 September 2024].

Demarchi, B. et al. (2020). Never boring: Non-invasive palaeoproteomics of mummified human skin. *Journal of archaeological science*, 119, Elsevier., p.105145. [Online]. Available at: doi:10.1016/j.jas.2020.105145.

Dierickx, K. et al. (2022). Peptide mass fingerprinting of preserved collagen in archaeological fish bones for the identification of flatfish in European waters. *Royal Society open science*, 9 (7), royalsocietypublishing.org., p.220149. [Online]. Available at: doi:10.1098/rsos.220149.

Dimoula, A. et al. (2019). Experimental investigation of ceramic technology and plant food cooking in Neolithic northern Greece. *STAR: Science & Technology of Archaeological Research*, 5 (2), Routledge., pp.269–286. [Online]. Available at: doi:10.1080/20548923.2020.1762370.

Earth Biogenome Project. (2022). *Earth Biogenome Project*. [Online]. Available at: <https://www.earthbiogenome.org/> [Accessed 19 September 2024].

East, C. P. et al. (2011). Deposition of hydroxyapatite and calcium oxalate dihydrate on a heat exchanger tube. *Asia-Pacific journal of chemical engineering*, 6 (6), Wiley., pp.921–932. [Online]. Available at: doi:10.1002/apj.487.

Engels, I. et al. (2024). ClassiCOL: LC-MS/MS analysis for ancient species Classification via Collagen peptide ambiguation. *bioRxiv*, p.2024.10.01.616034. [Online]. Available at: doi:10.1101/2024.10.01.616034 [Accessed 3 October 2024].

Enghoff, I. (2011). *Regionality and biotope exploitation in Danish Ertebølle and adjoining periods*, Scientia Danica. Series B Biologica. Det Kongelige Danske Videnskabernes Selskab.

ERGA. *A genome atlas of european biodiversity*. [Online]. Available at: <https://www.erga-biodiversity.eu/> [Accessed 14 October 2024].

Evans, M. et al. (2023). Detection of dairy products from multiple taxa in Late Neolithic pottery from Poland: an integrated biomolecular approach. *Royal Society open science*, 10 (3), p.230124. [Online]. Available at: doi:10.1098/rsos.230124.

Evans, M. et al. (2024). The impact of cooking and burial on proteins: a characterisation of experimental foodcrusts and ceramics. *Royal Society open science*, 11 (240610). [Online]. Available at: doi:10.1098/rsos.240610.

Ferro, L. et al. (2019). Statistical methods for rapid quantification of proteins, lipids, and carbohydrates in Nordic microalgal species using ATR–FTIR spectroscopy. *Molecules (Basel, Switzerland)*, 24, mdpi.com. [Online]. Available at: doi:10.3390/molecules24183237.

Fonseca, B., Freeman, C. L. and Collins, M. J. (2024). The survival of β -lactoglobulin peptides in the archaeological record: Vulnerability vs. Sequence variation. *bioRxiv*, p.2024.09.13.612646. [Online]. Available at: doi:10.1101/2024.09.13.612646 [Accessed 17 September 2024].

Fornelli, L. et al. (2018). Top-down proteomics: Where we are, where we are going? *Journal of proteomics*, 175, pmc.ncbi.nlm.nih.gov., pp.3–4. [Online]. Available at: doi:10.1016/j.jprot.2017.02.002.

Ge, W. et al. (2021). Neolithic bone meal with acorn: analyses on crusts in pottery bowls from 7000 BP Hemudu, China. *International journal of osteoarchaeology*, (oa.3026), Wiley. [Online]. Available at: doi:10.1002/oa.3026.

Ge, W. et al. (2024). Beyond rice: Preliminary investigation of encrusted organic residues in a pottery vessel reveals diverse food resources at the Neolithic Hemudu Settlement, China. *Journal of archaeological science, reports*, 58 (104714), Elsevier BV., p.104714. [Online]. Available at: doi:10.1016/j.jasrep.2024.104714.

Gianvincenzo, F. D. et al. (2023). Proteomic identification of beer brewing products in the ground layer of Danish Golden Age paintings. *Science Advances*, 9 (21), p.eade7686. [Online]. Available at: doi:10.1126/sciadv.ade7686.

Gurdeep Singh, R. et al. (2019). Unipept 4.0: Functional Analysis of Metaproteome Data. *Journal of proteome research*, 18 (2), pp.606–615. [Online]. Available at: doi:10.1021/acs.jproteome.8b00716.

Heiss, A. G. et al. (2020). Mashers to Mashers, Crust to Crust. Presenting a novel microstructural marker for malting in the archaeological record. *PloS one*, 15 (5), p.e0231696. [Online]. Available at: doi:10.1371/journal.pone.0231696.

Hendy, J. et al. (2018a). A guide to ancient protein studies. *Nature ecology & evolution*, 2 (5), nature.com., pp.791–799. [Online]. Available at: doi:10.1038/s41559-018-0510-x.

Hendy, J. et al. (2018b). Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nature communications*, 9 (1), p.4064. [Online]. Available at: doi:10.1038/s41467-018-06335-6.

Heron, C. and Craig, O. (2015). Aquatic Resources in Foodcrusts: Identification and Implication. *Radiocarbon*, 57 (4), Cambridge University Press., pp.707–719. [Online]. Available at: doi:10.2458/azu_rc.57.18454 [Accessed 12 May 2022].

Holst, D. (2010). Hazelnut economy of early Holocene hunter–gatherers: a case study from Mesolithic Duvensee, northern Germany. *Journal of archaeological science*, 37 (11), Elsevier BV., pp.2871–2880. [Online]. Available at: doi:10.1016/j.jas.2010.06.028 [Accessed 23 July 2024].

Hu, L. et al. (2017). Proteomic study of the effect of different cooking methods on protein oxidation in fish fillets. *RSC advances*, 7 (44), The Royal Society of Chemistry., pp.27496–27505. [Online]. Available at: doi:10.1039/C7RA03408C [Accessed 12 November 2021].

Jeong, C. et al. (2018). Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (48), National Acad Sciences., pp.E11248–E11255. [Online]. Available at: doi:10.1073/pnas.1813608115.

Jeong, K., Kim, S. and Bandeira, N. (2012). False discovery rates in spectral identification. *BMC bioinformatics*, 13 Suppl 16 (S16), Springer Science and Business Media LLC., p.S2. [Online]. Available at: doi:10.1186/1471-2105-13-S16-S2 [Accessed 16 October 2024].

Koenig, C. et al. (2024). Automated high-throughput biological sex identification from archaeological human dental enamel using targeted proteomics. *bioRxiv*, p.2024.02.20.581140. [Online]. Available at: doi:10.1101/2024.02.20.581140 [Accessed 26 February 2024].

Kong, A. T. et al. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature methods*, 14 (5), Springer Science and Business Media LLC., pp.513–520. [Online]. Available at: doi:10.1038/nmeth.4256.

Kubiak-Martens, L. (1999). The plant food component of the diet at the late Mesolithic (Ertebølle) settlement at Tybrind Vig, Denmark. *Vegetation history and archaeobotany*, 8 (1-2), Springer Science and Business Media LLC., pp.117–127. [Online]. Available at: doi:10.1007/bf02042850.

Kubiak-Martens, L. (2002). New evidence for the use of root foods in pre-agrarian subsistence recovered from the late Mesolithic site at Halsskov, Denmark. *Vegetation history and archaeobotany*, 11 (1-2), Springer Science and Business Media LLC., pp.23–32. [Online]. Available at: doi:10.1007/s003340200003.

Kubiak-Martens, L. et al. (2024). Transformation of cereal grains: Botanical and chemical analysis of food residues encrusted on pottery from the Funnel Beaker settlement of Oldenburg LA 77, northern Germany. *PloS one*, 19 (1), p.e0296986. [Online]. Available at: doi:10.1371/journal.pone.0296986.

Li, H. et al. (2016). Evaluating the effect of database inflation in proteogenomic search on sensitive and reliable peptide identification. *BMC genomics*, 17 (Suppl 13), BioMed Central., p.1031. [Online]. Available at: doi:10.1186/s12864-016-3327-5 [Accessed 2 November 2024].

Lucquin, A. et al. (2023). The impact of farming on prehistoric culinary practices throughout Northern Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 120 (43), p.e2310138120. [Online]. Available at: doi:10.1073/pnas.2310138120.

Lyu, N. et al. (2024). Characterization of Pottery Foodcrusts Through Lipid and Proteomic Analyses: A Case Study from the Xiawan Site in Yixing City, East China. *Journal of archaeological science*, 161 (105902), papers.ssrn.com. [Online]. Available at: doi:10.1016/j.jas.2023.105902 [Accessed 8 September 2023].

Mackie, M. et al. (2018). Palaeoproteomic Profiling of Conservation Layers on a 14th Century Italian Wall Painting. *Angewandte Chemie*, 57 (25), pp.7369–7374. [Online]. Available at: doi:10.1002/anie.201713020.

Marcilla, A. et al. (2011). Analytical pyrolysis as a method to characterize tannery wastes. *Industrial & engineering chemistry research*, 50 (15), American Chemical Society (ACS)., pp.8994–9002. [Online]. Available at: doi:10.1021/ie200582k [Accessed 11 October 2024].

Mooradian, A. D. et al. (2020). ProteoClade: A taxonomic toolkit for multi-species and metaproteomic analysis. *PLoS computational biology*, 16 (3), p.e1007741. [Online]. Available at: doi:10.1371/journal.pcbi.1007741.

Muth, T. and Renard, B. Y. (2018). Evaluating de novo sequencing in proteomics: already an accurate alternative to database-driven peptide identification? *Briefings in bioinformatics*, 19 (5), Oxford University Press (OUP)., pp.954–970. [Online]. Available at: doi:10.1093/bib/bbx033 [Accessed 15 October 2024].

Nitsch, E. K., Charles, M. and Bogaard, A. (2015). Calculating a statistically robust $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ offset for charred cereal and pulse seeds. *STAR: Science & Technology of Archaeological Research*, 1 (1), Routledge., pp.1–8. [Online]. Available at: doi:10.1179/2054892315Y.0000000001.

Noble, W. S. (2015). Mass spectrometrists should search only for peptides they care about. *Nature methods*, 12 (7), Springer Science and Business Media LLC., pp.605–608. [Online]. Available at: doi:10.1038/nmeth.3450 [Accessed 16 October 2024].

Oldenburg, T. et al. (2021). A novel route for identifying starch diagenetic products in the archaeological record. *PloS one*, 16 (11), p.e0258779. [Online]. Available at: doi:10.1371/journal.pone.0258779.

Ostrom, P. H. et al. (2006). Unraveling the sequence and structure of the protein osteocalcin from a 42ka fossil horse. *Geochimica et cosmochimica acta*, 70 (8), Elsevier BV.,

pp.2034–2044. [Online]. Available at: doi:10.1016/j.gca.2006.01.004 [Accessed 22 October 2024].

Paladugu, R. et al. (2023). Your horse is a donkey! Identifying domesticated equids from Western Iberia using collagen fingerprinting. *Journal of archaeological science*, 149, p.105696. [Online]. Available at: doi:10.1016/j.jas.2022.105696.

Papakosta, V., Oras, E. and Isaksson, S. (2019). Early pottery use across the Baltic—a comparative lipid residue study on Ertebølle and Narva ceramics from coastal hunter-gatherer sites in southern Scandinavia, northern Germany and Estonia. *Journal of archaeological science*, 24, Elsevier., pp.142–151. [Online]. Available at: doi:10.1016/j.jasrep.2019.01.003.

Perkins, D. N. et al. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20 (18), Wiley., pp.3551–3567. [Online]. Available at: doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

Peters, C. et al. (2021). Species identification of Australian marsupials using collagen fingerprinting. *Royal Society open science*, 8 (10), royalsocietypublishing.org., p.211229. [Online]. Available at: doi:10.1098/rsos.211229.

Philippsen, B. (2012). *Variability of freshwater reservoir effects, Implications for radiocarbon dating of prehistoric pottery and organisms from estuarine environments*. Heinemeier, J. and Olsen, J. (Eds). PhD, Aarhus University.

Pipart, J. et al. (2024). MultiStageSearch: a multi-step proteogenomic workflow for taxonomic identification of viral proteome samples addressing database bias. *bioRxiv*. [Online]. Available at: doi:10.1101/2024.05.15.594287.

Price, D. T. et al. (2001). Smakkerup Huse: A Mesolithic Settlement in NW Zealand, Denmark. *Journal of Field Archaeology*, 28 (1-2), Routledge., pp.45–67. [Online]. Available at: doi:10.1179/jfa.2001.28.1-2.45.

Price, T. D. et al. (2018). Asnæs Havnepark: a late Mesolithic Ertebølle coastal site in western Sjælland, Denmark. *Danish Journal of Archaeology*, 7 (2), Routledge., pp.255–276. [Online]. Available at: doi:10.1080/21662282.2018.1551979.

Ramsøe, A. et al. (2020). DeamiDATE 1.0: Site-specific deamidation as a tool to assess authenticity of members of ancient proteomes. *Journal of archaeological science*, 115, p.105080. [Online]. Available at: doi:10.1016/j.jas.2020.105080.

Ramsøe, A. et al. (2021). Assessing the degradation of ancient milk proteins through site-specific deamidation patterns. *Scientific reports*, 11 (1), p.7795. [Online]. Available at: doi:10.1038/s41598-021-87125-x.

Richter, J. and Noe-Nygaard, N. (2003). A Late Mesolithic Hunting Station at Agernæs, Fyn, Denmark: Differentiation and Specialization in the late Ertebølle-Culture, heralding the Introduction of Agriculture? *Acta Archaeologica*, 74 (1), Brill., pp.1–32. [Online]. Available at: doi:10.1111/j.0065-001X.2003.aar740101.x [Accessed 13 December 2022].

Robinson, D. E. and Harild, J. A. (2009). Archaeobotany of an early Ertebølle (late Mesolithic) site at Halsskov, Zealand, Denmark. In: Mason, S. L. R. and Hather, J. (Eds). *Hunter-Gatherer Archaeobotany: Perspectives from the Northern Temperate Zone*. London, England : Routledge. [Online]. Available at: <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315427171-8/archaeobotany-early->

ertebe% C3%B8lle-late-mesolithic-site-halsskov-zealand-denmark-david-earle-robinson-jan-andreas-harild.

Rodriguez Palomo, I. et al. (2024). Benchmarking the identification of a single degraded protein to explore optimal search strategies for ancient proteins. *Peer Community J.*, (e107), p.2023.12.15.571577. [Online]. Available at: doi:10.24072/pcjournal.491 [Accessed 29 November 2024].

Rosmaninho, R. et al. (2007). Calcium phosphate fouling on TiN-coated stainless steel surfaces: Role of ions and particles. *Chemical engineering science*, 62 (14), Elsevier BV., pp.3821–3831. [Online]. Available at: doi:10.1016/j.ces.2007.04.008.

Sayers, E. W. et al. (2022). Database resources of the national center for biotechnology information. *Nucleic acids research*, 50 (D1), Oxford University Press (OUP)., pp.D20–D26. [Online]. Available at: doi:10.1093/nar/gkab1112 [Accessed 30 October 2024].

Shevchenko, A. et al. (2018). Fine Endmesolithic fish caviar meal discovered by proteomics in foodcrusts from archaeological site Friesack 4 (Brandenburg, Germany). *PloS one*, 13 (11), p.e0206483. [Online]. Available at: doi:10.1371/journal.pone.0206483.

Siano, F. et al. (2022). Proteomics and Integrated Techniques to Characterize Organic Residues in Funerary Findings from Italic Populations of the First Millennium BC. *Journal of proteome research*, 21 (5), ACS Publications., pp.1330–1339. [Online]. Available at: doi:10.1021/acs.jproteome.2c00093.

Smith, L. M. and Kelleher, N. (2013). Proteoform: a single term describing protein complexity. *Nature methods*, 10, nature.com., pp.186–187. [Online]. Available at: doi:10.1038/nmeth.2369.

Sørensen, L. and Karg, S. (2014). The expansion of agrarian societies towards the north – new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. *Journal of archaeological science*, 51, Elsevier BV., pp.98–114. [Online]. Available at: doi:10.1016/j.jas.2012.08.042.

Styring, A. K. et al. (2013). The effect of charring and burial on the biochemical composition of cereal grains: investigating the integrity of archaeological plant material. *Journal of archaeological science*, 40 (12), Elsevier BV., pp.4767–4779. [Online]. Available at: doi:10.1016/j.jas.2013.03.024.

Tanasi, D. et al. (2021). Paleoproteomic profiling of organic residues on prehistoric pottery from Malta. *Amino acids*, 53 (2), pp.295–312. [Online]. Available at: doi:10.1007/s00726-021-02946-4.

Tang, L. et al. (2023). Paleoproteomic evidence reveals dairying supported prehistoric occupation of the highland Tibetan Plateau. *Science advances*, 9 (15), p.eadf0345. [Online]. Available at: doi:10.1126/sciadv.adf0345.

Trolle, T. (2013). Human and animal bones from Tybrind Vig, Analysis of material from a burial and settlement. In: Andersen, S. H. (Ed). *Tybrind Vig, Submerged Mesolithic settlements in Denmark*. 77. Højbjerg : Jutland Archaeological Society. pp.413–478.

Tyanova, S., Temu, T. and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols*, 11 (12), nature.com., pp.2301–2319. [Online]. Available at: doi:10.1038/nprot.2016.136.

Uhlén, M. et al. (2019). The human secretome. *Science signaling*, 12 (609). [Online].

Available at: doi:10.1126/scisignal.aaz0274.

UniProt Consortium. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic acids research*, 51 (D1), pp.D523–D531. [Online]. Available at: doi:10.1093/nar/gkac1052.

Vieugué, J. et al. (2015). The Consumption of Bone Powder in the Early Neolithic Societies of Southeastern Europe: Evidence of a Diet Stress? *Cambridge Archaeological Journal*, 25 (2), Cambridge University Press., pp.495–511. [Online]. Available at: doi:10.1017/S0959774314001048 [Accessed 16 November 2021].

Wang, H.-Y., Qian, H. and Yao, W.-R. (2011). Melanoidins produced by the Maillard reaction: Structure and biological activity. *Food chemistry*, 128 (3), pp.573–584. [Online]. Available at: doi:10.1016/j.foodchem.2011.03.075.

Wang, S. et al. (2020). Starch-lipid and starch-lipid-protein complexes: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*, 19 (3), Wiley., pp.1056–1079. [Online]. Available at: doi:10.1111/1541-4337.12550.

Warinner, C. et al. (2014). Direct evidence of milk consumption from ancient human dental calculus. *Scientific reports*, 4, nature.com., p.7104. [Online]. Available at: doi:10.1038/srep07104.

Warinner, C., Korzow Richter, K. and Collins, M. J. (2022). Paleoproteomics. *Chemical reviews*, ACS Publications. [Online]. Available at: doi:10.1021/acs.chemrev.1c00703.

Welker, F. et al. (2020). The dental proteome of Homo antecessor. *Nature*, 580 (7802), pp.235–238. [Online]. Available at: doi:10.1038/s41586-020-2153-8.

Wilkin, S. et al. (2021). Dairying enabled Early Bronze Age Yamnaya steppe expansions. *Nature*, 598 (7882), Springer Science and Business Media LLC., pp.629–633. [Online]. Available at: doi:10.1038/s41586-021-03798-4 [Accessed 9 September 2024].

Wilkin, S. et al. (2023). Curated cauldrons: Preserved proteins from early copper-alloy vessels illuminate feasting practices in the Caucasian steppe. *iScience*, 26 (9), p.107482. [Online]. Available at: doi:10.1016/j.isci.2023.107482.

Yilmaz, M. et al. (2022). De novo mass spectrometry peptide sequencing with a transformer model. In: *International Conference on Machine Learning*. 28 June 2022. PMLR. pp.25514–25522. [Online]. Available at: <https://proceedings.mlr.press/v162/yilmaz22a.html> [Accessed 6 September 2024].

Yilmaz, M. et al. (2024). Sequence-to-sequence translation from mass spectra to peptides with a transformer model. *Nature communications*, 15 (1), Springer Science and Business Media LLC., p.6427. [Online]. Available at: doi:10.1038/s41467-024-49731-x [Accessed 6 September 2024].

Yu, T.-Y. et al. (2017). Cooking-induced protein modifications in meat. *Comprehensive Reviews in Food Science and Food Safety*, 16 (1), Wiley., pp.141–159. [Online]. Available at: doi:10.1111/1541-4337.12243.

Zhang, J. et al. (2021). Fouling behavior of calcium phosphate in direct contact membrane distillation. *Environmental technology & innovation*, 21 (101203), Elsevier BV., p.101203. [Online]. Available at: doi:10.1016/j.eti.2020.101203 [Accessed 10 October 2024].

Zheng, M. et al. (2018). Effects of chain length and degree of unsaturation of fatty acids on structure and in vitro digestibility of starch-protein-fatty acid complexes. *Journal of*

agricultural and food chemistry, 66 (8), ACS Publications., pp.1872–1880. [Online]. Available at: doi:10.1021/acs.jafc.7b04779.

Chiang, Y. (2024). Unpublished concept version of the MS/MS identification pipeline Cerberus provided by Yun Chiang, 22 July 2024.

7 Conclusion

Foodcrusts hold a lot of potential for improving our understanding of culinary practices in the past. They provide direct evidence of the cooking of foods and can yield evidence of both the animal and plant component of the diet. Yet, simultaneously they are a challenging material to analyse. The heating necessary to produce foodcrusts damages and modifies proteins. As a result, the number of proteins recovered from foodcrust is typically relatively low. Additionally, many aspects of the composition of foodcrust and their formation remain unknown. This thesis aimed to contribute to our understanding of foodcrust as a material type for dietary analysis by investigating the formation of foodcrust and how recovered proteins relate to the foodstuffs that formed the foodcrust.

To highlight the magnitude of the differences in the ease of extracting proteins from foodcrust compared to the more commonly analysed biomineral tissues a bone tool from the Bronze Age of Heiloo, the Netherlands, was analysed with ZooMS, a SPIN-like bioinformatic workflow and a targeted database MaxQuant search. The ZooMS analysis indicated that the sample likely was from either *Eubalaena glacialis* or *Balaena mysticetus*, while the SPIN-like analysis indicated *Balaenoptera acutorostrata*, because the SPIN database contained no entries for *Eubalaena glacialis* or *Balaena mysticetus*. Finally, by employing a custom made database of Mysticeti collagen sequences, a MaxQuant search confirmed that the recovered proteins only matched to *Eubalaena glacialis*, the northern right whale, showing that it is possible to distinguish between closely related species using palaeoproteomics.

Species-specific protein identification might be possible in some cases, but not all foodstuffs that are likely to have been cooked in the past are equally protein-rich. In order to test how fairly the ingredient proportions of a mixture are represented by the recoverable proteins from the foodcrust and how this is impacted by additional cooking events and burial, a protein-rich (salmon), lipid-rich (beeffat) and carbohydrate-rich (beetroot) ingredient were mixed at different proportions, cooked nine times for one hour at 270°C and buried for 52 days. The change in protein recovery compared to the bone tool of Chapter 2 is stark as 25 out of 84 samples yielded no proteins at all. Protein recovery decreased significantly with increased cooking events and burial. Of the 21 samples taken after burial only 7 still yielded proteins. Additionally, no beef proteins could be observed after the burial, while most of the remaining proteins were salmon. Across the entire dataset, proteins from all cooked species could be identified in only 32 of 84 samples. The results also indicated large variations in the amount of foodcrust generated by combining the same three ingredients in different proportions. No single clear predictor of foodcrust volume could be identified, but the results suggest that carbohydrate- and protein-rich foods are more important than lipid-rich foodstuffs in forming foodcrust. The experiment showed not all ingredients of a cooked mixture will be equally represented in the foodcrust and that in general the amount of protein that can be recovered from a foodcrust will be low.

The blind test reported in Chapter 4 consisted of twenty foodcrust samples that were analysed using two different mass spectrometers (Fusion and timsTOF), two extraction protocols (SP3 and DES) and two different bioinformatic validation workflows (Unipept and ProteoClade). First of all, results showed that the timsTOF managed to identify more proteins than the Fusion analysis, but the difference between the two extraction protocols is more complex. The SP3 method extracted proteins from a higher number of samples than the DES method, but the DES method often extracted more proteins per sample. In terms of

the accuracy of the taxonomic identification there were only three samples among the twenty of which all ingredients could be accurately identified following conventional identification criteria. From all other samples either one of the ingredients could not be recovered or false positives were detected. For some of the species it is clear that they could not be identified due to a lack of available reference proteins, but other samples contained foodstuffs that have been identified in archaeological samples before, such as dairy. Furthermore, only by increasing the minimum number of associated peptides to six and summarising all taxonomic identifications to order level could most of the false positives be removed. The blind test highlights that the conventional rule of two peptides per protein is in some cases insufficient to weed out incorrect identifications.

Lastly, the analysis of the Mesolithic and Neolithic samples from Syltholm II showed substantial differences between the different methods. SEM analysis identified wild plant taxa in both the Mesolithic and Neolithic vessels, such as *Chenopodiaceae* and *Viburnum opulus*, as well as likely fish bone fragments. Additionally, a cereal fragment cf. *Hordeum vulgare* was found in a Neolithic foodcrust. The lipid analysis revealed great similarities between the Mesolithic and Neolithic samples, both showing evidence of dairy, marine resource exploitation and likely mixtures of the two. Only a small number of samples yielded proteins, but dairy proteins were identified in both the Mesolithic and Neolithic samples supporting the notion that there was trade in dairy between Mesolithic hunter gatherers and Neolithic farmers in Denmark at the start of the fourth millennium. The identified dairy proteins were all alpha-s1-casein and could be specified to *Bos taurus*. No beta-lactoglobulin was found, which might mean the dairy proteins originated not from milk, but a curd product. Additionally one of the Neolithic samples contained *Pisum sativum* and frog proteins, but it is not completely sure if these are authentic ancient proteins.

The overarching conclusion of this thesis is that the analysis of proteins from foodcrust can yield valuable information, but is more challenging and should be performed with more caution than current conventional practices. The amount of recoverable protein from foodcrust is limited and there is a substantial risk that no proteins can be recovered from a sample at all. Additionally, the species-level taxonomic resolution achievable with bone samples is seldom possible with foodcrusts and there is a risk of false positives that cannot be neglected. However, when proteins are recovered they might provide valuable new insights, such as identifying the presence of dairy in Mesolithic Denmark.