Development and application of palaeoproteomics (ZooMS) on Palaeolithic assemblages from Europe and East Asia

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Naihui Wang

aus Jinan, China

Tübingen

2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	03.06.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Nicholas J. Conard
2. Berichterstatter:	Assoc. Prof. Katerina Douka
3. Berichterstatter:	Jun. Prof. Cosimo Posth

Acknowledgements

Thirteen years ago, as I sat in my high school physics classroom contemplating the time it takes for charged particles to traverse electromagnetic fields, the teacher mentioned, "Some of you may find yourselves using this equipment in the future." At that moment, I couldn't fathom that one of those "some" would be none other than myself.

I am profoundly grateful to my supervisor, Katerina Douka, for guiding me into the enchanting realm of ZooMS. Since then, the exploration of the flight of charged peptides within mass spectrometers has become one of the most captivating journeys of my life. Throughout my five-year doctoral research, Katerina provided steadfast support, even accommodating my shortcomings in English writing. Words fail to express the depth of my gratitude to Katerina, regardless of the language used.

I also extend my thanks to my supervisor, Nicholas Conard. It is under your guidance that I entered the Palaeolithic caves in Swabian Jura, making this representative of Generation Z humble in the face of the artistic creations of Swabian Aurignicians. Your generous academic guidance and passion for research always encouraged me, and the archaeological excavation opportunities you provided allow me to proudly identify as an archaeologist.

Special thanks to Samantha Brown, Annette Oertle, Kristine Korzow Richter and Noel Amano for their training and help, enabling me to complete this research successfully.

I appreciate my Chinese collaborators Yang Xu, Yinqiu Cui, Zhuowei Tang, Cunding He, Gang Li, Gaike Zhang, Xiaomin Wang and Yue Zhang, who provided samples and laboratory support, especially during the challenging years of the world pandemic. Without you, the completion of the first large-scale ZooMS project in East Asia would have been impossible.

Gratitude is extended to the anonymous reviewers of the journals, your insights significantly enhanced the quality of my articles every time.

To my lifelong friends Yun Zhang, Yufei Zhou, Ke Wang, and Lei, your companionship across time zones has been invaluable in preventing feelings of loneliness.

I thank my co-windows Li Tang, Traci Billings, Barbara Huber, Martha Kayuni, and Eliane Chim. In Chinese, "co-windows" refers to study partners under the same window. I am honoured to have shared an office with you, and I hope our PhD research all progresses smoothly.

Thanks to the support from the European Research Council (ERC) on FINDER project, awarded to my supervisor Katerina Douka (715069-FINDER-ERC-2016-STG), for funding my doctoral research.

Table of Contents

List of Figures

Abbreviations

ACN:	acetonitrile
aDNA:	ancient desoxyribonucleic acid
AH:	archaeological horizon
AmBic:	ammonium bicarbonate
AMS:	accelerator mass spectrometry
ATR-FTIR:	attenuated total reflectance Fourier transform infrared spectroscopy
C18:	octyldecylsilane
CAA:	chloroacetamide
cal ka BP:	calibrated thousand years before present
CHCA:	α -cyano-4-hydroxycinnamic acid
COL1:	type I collagen
COL1a1:	alpha 1 chain of type I collagen
COL1a2:	alpha 2 chain of type I collagen
EDTA:	ethylendiaminetretraacetic acid
ESR:	electron spin resonance dating
EtOH:	ethanol
HCl:	hydrogen chloride
HPLC:	high-performance liquid chromatography
Int:	mass intensity
kDa:	thousand daltons
LC-MS/MS:	liquid chromatography with tandem mass spectrometry
m/z:	mass-to-charge ratio
MALDI-FTICR-MS:	matrix-assisted laser desorption Fourier transform ion cyclotron resonance mass spectrometry
MALDI-TOF-MS:	matrix-assisted laser desorption ionization time-of-flight mass spectrometry
MALDI-TOF/TOF-MS:	matrix-assisted laser desorption ionization time-of-flight tandem mass spectrometry
MNI:	minimum number of individuals
%N:	percent nitrogen by weight
NH_4HCO_3:	ammonium bicarbonate

NISP:	number of identified specimens
OSL:	optically stimulated luminescence dating
PTM:	post-translational modification
Q:	glutamine amino acid
s/n:	signal-to-noise ratio
SP3:	single-pot solid-phase-enhance sample preparation
TCEP:	TCEP: tris (2-carboxyethyl) phosphine
TFA:	TFA: trifluoroacetic acid
ZooMS:	zooarchaeology by mass spectrometry

ZooMS markers (Brown et al., 2021; Welker et al., 2016)

P1:	COL1a1 507 - 518
P2:	COL1a2 292 - 309
A:	COL1a2 978 - 990
B:	COL1a2 484 - 498
C:	COL1a2 502 - 519
D:	COL1a2 793 - 816
E:	COL1a2 454 - 483
F:	COL1a1 585 - 617
G:	COL1a2 757 - 789

Summary

Over the past decade, the broader application of palaeoproteomics, particularly Zooarchaeology by Mass Spectrometry (ZooMS), has led to the discovery of new hominin fossils across various Palaeolithic sites. ZooMS, with its capacity to taxonomically identify faunal remains based on molecular features, has significantly expanded the scope of faunal research of the Pleistocene. Nonetheless, the workflow of ZooMS analysis requires further refinement, particularly in terms of chemical protocol applicability and data processing. It is also notable that the application of ZooMS has yet to be extended to East Asia.

This doctoral research pursued three main objectives. First, I aimed to assess the performance of various ZooMS protocols for screening projects and subsequently to determine the optimal protocol for samples of diverse preservation. Thus, I evaluated three existing ZooMS protocols and a new SP3 protocol. Through the analysis of 400 samples from seven sites, I concluded that the Acid-insoluble protocol consistently yielded optimal collagen extracts, while the non-destructive AmBic protocol proved effective only for well-preserved samples. The newly tested SP3 protocol exhibited inferior performance compared to other protocols.

The second objective was to enhance the efficiency of data processing for screening large numbers of bones using ZooMS. I introduced a semi-automated identification pipeline for ZooMS (QuickID-based) and tested it using a manually-identified dataset. Results demonstrated that QuickID accelerates data processing, ensuring consistent standards of identification and improving overall reproducibility.

The third objective involved applying the refined ZooMS workflow to Palaeolithic fauna collections from Europe and East Asia, then integrating ZooMS results and traditional zooarchaeological data. In the first ZooMS screening project, I analysed 287 bone fragments from Vogelherd Cave in the Swabian Jura, which provided insights regarding the undiagnostic small bones and resulted in the discovery of three new human bones. In the second project, I analysed 745 bones from Jinsitai Cave, an important Middle and Upper Palaeolithic site on the Mongolia Plateau in China. I uncovered two taxa - camel and ostrich - that had not previously been identified in the zooarchaeological study. The subsequent attempts on integration established a

framework for combining ZooMS and zooarchaeological data in Palaeolithic contexts, aiding analysts in formulating research questions and making informed decisions regarding the use of ZooMS.

Additional ZooMS analyses of nearly 600 bones from four Chinese Palaeolithic sites yielded exciting insights into understanding how sedimentary context contributes to collagen preservation. More importantly, the results confirmed that ZooMS could be applied to very old (>100 ka) contexts with high success rates in some favourable conditions in southern China.

This dissertation encompasses refined analytical workflows for ZooMS, successful applications in European and East Asian Palaeolithic contexts, and insights into the integration of ZooMS and traditional zooarchaeology. These results have the potential to propel the field of palaeoproteomics forward and serve as a useful resource for future studies.

Zusammenfassung

Im Laufe des letzten Jahrzehnts hat die breitere Anwendung der Paläoproteomik, insbesondere der Zooarchäologie mittels Massenspektrometrie (ZooMS), zur Entdeckung neuer Hominiden-Fossilien an verschiedenen paläolithischen Fundorten geführt. ZooMS hat dadurch, dass faunische Überreste auf der Grundlage molekularer Merkmale taxonomisch identifiziert werden können, den Umfang der der Forschung zur Fauna im Pleistozän erheblich erweitert. Dennoch bedarf die ZooMS-Analyse weiterer Verfeinerungen, insbesondere hinsichtlich der Anwendbarkeit von chemischen Protokollen und der Datenverarbeitung. Außerdem wurde die Anwendung von ZooMS bisher nicht auf Ostasien ausgedehnt.

Diese Doktorarbeit verfolgte drei Hauptziele. Erstens sollte die Leistungsfähigkeit verschiedener ZooMS-Protokolle für Großprojekte bewertet und anschließend das optimale Protokoll für Proben mit unterschiedlichen Konservierungen bestimmt werden. Daher habe ich drei vorhandene ZooMS-Chemieprotokolle und ein neues SP3-Protokoll evaluiert. Durch die Analyse von 400 Proben von sieben Standorten kam ich zu dem Schluss, dass das säureunlösliche Protokoll durchweg optimale Kollagenextrakte lieferte, während das nicht-destruktive AmBic-Protokoll nur für gut erhaltene Proben effektiv war. Das neu getestete SP3-Protokoll zeigte eine geringere Leistung im Vergleich zu anderen Protokollen.

Das zweite Ziel bestand darin, die Effizienz der Datenverarbeitung für Großprojekte mit ZooMS-Anwendungen zu verbessern. Ich führte eine teilautomatisierte Identifikationspipeline für ZooMS (QuickID-basiert) ein und testete sie anhand eines manuell identifizierten Datensatzes. Die Ergebnisse zeigten, dass QuickID die Datenverarbeitung beschleunigt, konsistente Identifikationsstandards sicherstellt und die Gesamtreproduzierbarkeit verbessert.

Das letzte Ziel bestand darin, den verfeinerten ZooMS-Arbeitsablauf auf paläolithische Faunensammlungen aus Europa und Ostasien anzuwenden und die Ergebnisse mit traditionellen zooarchäologischen Daten zu kombinieren. Im ersten ZooMS-Screening-Projekt analysierte ich 287 Knochenfragmente aus der Vogelherdhöhle in der Schwäbischen Alb, was Erkenntnisse zu faunischen Überresten kleiner Größe lieferte und zur Entdeckung von drei neuen menschlichen Knochen führte. Im zweiten

Projekt analysierte ich 745 Knochen aus der Jinsitai-Höhle, einem wichtigen mittel- und jungpaläolithischen Fundort auf dem Mongolischen Plateau in China. Ich entdeckte zwei neue Taxa – Kamel und Strauß –, die zuvor nicht in der zooarchäologischen Gemeinschaft identifiziert worden waren. Die anschließenden Untersuchungen an beiden Standorten schufen einen Rahmen für die Kombination von ZooMS- und zooarchäologischen Daten in paläolithischen Kontexten, was Analysten bei der Formulierung von Forschungsfragen und fundierten Entscheidungen über die Verwendung von ZooMS unterstützt.

Zusätzliche ZooMS-Analysen von fast 600 Knochen aus vier chinesischen paläolithischen Fundorten ergaben interessante Erkenntnisse darüber, wie der sedimentäre Kontext zur Kollagenerhaltung beiträgt. Noch wichtiger ist, dass die Ergebnisse bestätigten, dass ZooMS unter günstigen Bedingungen in Südchina mit hohen Erfolgsquoten auf sehr alte (>100 ka) Kontexte angewendet werden kann.

Diese Dissertation umfasst verfeinerte analytische Workflows für ZooMS, erfolgreiche Anwendungen in europäischen und ostasiatischen paläolithischen Kontexten und Einblicke in die Kombination von ZooMS mit traditioneller Zooarchäologie. Die Ergebnisse haben das Potenzial, das Feld der Paläoproteomik voranzutreiben und diese Arbeit als nützliche Ressource für zukünftige Studien zu etablieren.

List of publications

Accepted manuscripts (Appendix 1)

Manuscript A - Naihui Wang, Samantha Brown, Peter Ditchfield, Sandra Hebestreit, Maxim Kozilikin, Sindy Luu, Oshan Wedage, Stefano Grimaldi, Michael Chazan, Liora Kolska Horwitz, Matthew Spriggs, Glenn Summerhayes, Michael Shunkov, Kristine Korzow Richter, Katerina Douka. "Testing the efficacy and comparability of ZooMS protocols on archaeological bone." *Journal of Proteomics* 233 (2021): 104078.

Manuscript B - Naihui Wang, Xu Yang, Zhuowei Tang, Cunding He, Xin Hu, Yinqiu Cui, Katerina Douka. "Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China." *Proceedings of the Royal Society B* 290, no. 2009 (2023): 20231129.

Manuscript C - Naihui Wang, Nicholas J. Conard, Katerina Douka. “Integrating morphological and ZooMS-based approaches to zooarchaeology at Vogelherd Cave in Southwestern Germany”, *PaleoAnthropology*. Special Issue: Integrating ZooMS and Zooarchaeology: Methodological Challenges and Interpretive Potentials (2024).

Other publication (Appendix 2)

Yang Xu, Naihui Wang, Shizhu Gao, Chunxiang Li, Pengcheng Ma, Shasha Yang, Hai Jiang, Shoujin Shi, Yanhua Wu, Quanchao Zhang, Yinqiu Cui. "Solving the two-decades-old murder case through joint application of ZooMS and ancient DNA approaches." *International Journal of Legal Medicine* 137, no. 2 (2023): 319-327.

Personal contributions

Manuscript A - (Wang et al., 2021) K. Douka, S. Brown and N. Wang proposed and designed this project. M. Kozilikin, S. Luu, O. Wedage, S. Grimaldi, M. Chazan, L.K. Horwitz, M. Spriggs, G. Summerhayes, M. Shunkov provided samples. N. Wang, S. Brown and S. Hebestreit performed the ZooMS analysis of 400 bones at the ZooMS Laboratory, Max Planck Institute of Geoanthoropology. P. Ditchfield measured the nitrogen percentage of 400 bones at the Research Laboratory for Archaeology and the History of Art, School of Archaeology, University of Oxford. N. Wang analysed the data and created the figures and tables. N. Wang, K. Douka and K.K. Richter wrote the paper with the input and assistance of all authors.

In total, N. Wang contributed 75% to the project, including ZooMS lab work, data processing, visualisation and manuscript construction.

Manuscript B - (Wang et al., 2023) K. Douka, N. Wang and Y. Cui conceived this project. Z. Tang, C. He and X. Hu provided the samples. N. Wang and Y. Xu drilled 866 samples from Jinsitai and Yumidong, and performed ZooMS analysis at the Palaeoproteomics Laboratory, School of Life Science, Jilin University. N. Wang and K. Douka analysed the data, created the figures and tables. N. Wang and K. Douka wrote the paper, with input from Y. Xu.

In total, N. Wang contributed 80% to the project, including establishing the laboratory, sampling, ZooMS lab work, data visualisation, and manuscript construction.

Manuscript C - (Wang et al., 2024) N.J. Conard and K. Douka developed this project. N.J. Conard provided the samples. N. Wang drilled 287 Vogelherd samples and performed ZooMS analysis at the ZooMS Laboratory, Max Planck Institute of Geoanthoropology. N. Wang processed the data, created the figures and tables. N. Wang, N.J. Conard and K. Douka wrote the paper.

In total, N. Wang contributed 85% to the project, including ZooMS lab work, data processing, visualisation and manuscript construction.

1. Introduction

Zooarchaeology by mass spectrometry (ZooMS) (Buckley et al., 2009), a variant of peptide mass fingerprinting, has been employed in archaeology for over a decade. It is a rapid and cost-effective method for taxonomically identifying collagenous materials, particularly bones (Brandt et al., 2020; Brown et al., 2021c; Fiddyment et al., 2015). ZooMS entails the extraction of collagen from different materials and the generation of tryptic-digested peptide mass fingerprints using MALDI-TOF-MS (Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry).

Bones are the most prevalent collagenous materials in archaeological records. The bone extracellular matrix is a composite of the mineral fraction and the organic matrix. The former primarily comprises calcium phosphate or hydroxyapatite, while the organic matrix mainly consists of proteins, including collagen (90%) and other non-collagenous proteins (3 to 5%) (Weiner and Wagner, 1998). Due to the substantial collagen content in living bone and the intertwining of the mineral matrix with collagen fibrils, bone collagen exhibits great preservation properties (Dobberstein et al., 2009; Wadsworth and Buckley, 2014).

Bulk collagen has been extracted routinely from bones for stable isotope analysis or radiocarbon dating (Longin, 1971). In ZooMS analysis, the targeted molecule is Type I collagen (COL1), the predominant collagen in bone. COL1 is a fibrillar protein comprising two identical COL1a1 chains and one COL1a2 chain in tetrapods, each ~1400 amino acids in length, collectively forming a triple helix (Shoulders and Raines, 2009).

COL1 is essential for bone structure, and its sequence is highly conserved across taxa. Nonetheless, variations in amino acid sequences among different taxa enable the differentiation (Buckley, 2018). ZooMS circumvents the need for sequencing, relying instead on distinctions in molecular weights of various protein fragments. In the ZooMS protocol, collagen (largely COL1) is liberated from the bone's inorganic matrix by dissolution into a suitable buffer, amenable to enzymatic digestion, with or without demineralisation treatment. Subsequently, enzymes digest the extract into protein fragments known as peptides. Following certain purification steps, the resulting peptide is spotted onto a plate with a solvent absorbing laser energy, leading to co-

crystallization, followed by analysis via MALDI-TOF-MS (Buckley et al., 2010). Taxonomic identification hinges on diagnostic peptides evident in spectra, termed ZooMS markers. Over the past decade, the reference of ZooMS markers has expanded from extant species to extinct ones, spanning mammals, fishes, birds, and regions encompassing Europe, Africa, and Australia (Buckley et al., 2010; Codlin et al., 2022; Janzen et al., 2021; Peters et al., 2021; Richter et al., 2011; Welker et al., 2016).

Since its initial inception, ZooMS has been applied to a wide array of collagenous materials beyond bones in diverse contexts (e.g., leather, ivory, and parchment) (Brandt and Mannering, 2021; Coutu et al., 2016; Fiddyment et al., 2015). Nevertheless, this doctoral research is confined to applying ZooMS on bone assemblages from Palaeolithic sites.

Palaeolithic sites often yield highly fragmented bones due to human or carnivore-induced modifications and taphonomic processes. These non-diagnostic bones, lacking distinctive features, pose challenges on taxonomic identification through morphological approaches. Here, ZooMS presents a molecular avenue for identifying such bones. The application of ZooMS has been successful in Palaeolithic contexts, particularly for discovering new hominin fossils.

For instance, researchers uncovered at least nine additional hominin fossils from over 10,000 bone fragments at Denisova Cave using ZooMS (Brown et al., 2021b, 2016). Similar discoveries have occurred at several other European sites (Devièse et al., 2017; Hublin et al., 2020; Mylopotamitaki et al., 2024; Welker et al., 2016). Human remains are scarce in Palaeolithic contexts, and these novel finds, coupled with techniques like radiocarbon dating and aDNA analysis, contribute valuable insights into the evolutionary trajectory of our species (Slon et al., 2018). As reported by Rybczynski et al. (2013), ZooMS analysis can be extended to bones dating back as far as 3.5 million years in some parts of the world, encouraging the pursuit of ZooMS screening projects spanning a wide temporal range.

The scope of this doctoral research focuses on ZooMS screening projects, which typically involve high-throughput analyses of hundreds of specimens to uncover new human remains. It contrasts with the smaller-scale ZooMS analyses tailored to more specific zooarchaeological inquiries. Moreover, most materials in this research

originate from collections lacking or have lost stratigraphic information, either due to coarse excavation practices or extended storage periods. Furthermore, in the context of old collections, zooarchaeological investigations based on identifiable fauna assemblages are generally complete and published, which offer valuable references to ZooMS analysts.

This doctoral research is situated within the overarching framework of the FINDER project, funded by the European Research Council. FINDER aims to conduct comprehensive screening, analysis, and characterisation of newly discovered hominin fossils originating from Pleistocene Eurasia. Consequently, this research spans various archaeological sites, encompassing locations in both Europe and East Asia. Notably, the East Asian sites represent uncharted territory for ZooMS investigations, marking a previously unexplored dimension in our pursuit of knowledge.

1.1 Objectives of the dissertation and research questions

Several challenges underlie high-throughput ZooMS screening. In this dissertation, I endeavour to propose solutions pertaining to protocol selection, spectra identification, and data integration with zooarchaeological datasets. The objectives of this work are:

1.1.1 Testing protocol efficiency

To facilitate ZooMS analysis in more challenging environments where collagen preservation is not optimal, the selection of an appropriate protocol holds significant importance. A rich history of bone collagen research predates the advent of ZooMS. The traditional extraction protocol involves demineralisation using strong acid (~1M HCl) and gelatinisation of the insoluble pellet through gentle heating in weak acid (~0.001M HCl). The resulting gelatin is combusted for stable isotope or radiocarbon dating analyses (Longin, 1971). In the initial ZooMS protocol (Buckley et al., 2009), strong HCl is again employed for demineralisation, and then ammonium bicarbonate (AmBic) is used as the gelatinisation solvent to enable compatibility with subsequent trypsin digestion. In cases of limited sample quantities or low collagen yield from the insoluble fraction, the HCl soluble fraction obtained during demineralisation is combined for ZooMS (van der Sluis et al., 2014). Consideration of time efficiency also makes the HCl soluble fraction a preferred choice (Harvey et al., 2022). An alternative

non-destructive method, widely accepted for bone artefacts (van Doorn et al., 2011), avoids the demineralisation step. This protocol uses AmBic to gelatinise collagen (or collagen fragments) from the bone surface, mitigating damage to the bone's inorganic matrix. For well-preserved samples, this gentler protocol yields sufficient collagen for ZooMS (Martisius et al., 2020).

Concerns arise regarding poorly preserved samples since demineralisation by strong acid could lead to collagen hydrolysis (Collins and Galley, 1998). Additionally, the acid-based protocol involves multiple solvent-transferring steps, which inevitably leave some solvent in tubes and lead to a reduction in diagnostic peptide recovery for ZooMS identification. Thus, SP3 (single-pot solid-phase-enhance sample preparation), a bead-based protocol commonly employed in proteomic sample preparation (Hughes et al., 2014), can be a promising alternative to routine ZooMS protocols for poorly preserved samples. Performing entirely within a single tube, SP3 minimises protein loss during washing and transferring of the sample. Cleland (2018) introduced the SP3 application to ancient bone samples, with EDTA as the demineralisation buffer. EDTA can gently and gradually demineralise bone matrix by chelating calcium and releasing phosphate (Cleland et al., 2012; Collins and Galley, 1998).

The primary objective of this doctoral research was to assess the efficacy of the routine ZooMS protocols and SP3, by comparing protein extraction completeness, contaminant removal, and success rates for taxonomic identification across a wide range of samples with varying preservation status. An independent proxy for collagen preservation, the percent nitrogen by weight (%N), was also measured. A total of 400 bone samples from seven sites were examined, including sites from challenging tropical and savanna regions. (Manuscript A).

1.1.2 Batch data processing

The ZooMS fingerprint of a sample is visualised in a MALDI-TOF-MS spectrum, wherein each peak corresponds to a cluster of COL1 peptides sharing the same mass. Only several mass peaks can function as diagnostic markers in ZooMS analysis, indicating peptides with specific amino acid sequences and taxonomically informative. The ensemble of markers within a spectrum forms the COL1 fingerprint of a given sample.

Interpreting ZooMS spectra is a relatively straightforward process: analysts compare observed peaks with those of referenced species, document these diagnostic markers, and perform identification. This process is predominantly performed manually in most ZooMS studies, often following a workflow similar to that outlined by Brown (2021).

Manually identification of mass spectra can be time-consuming, particularly when working with extensive datasets. To address this challenge, researchers have developed some automated approaches for ZooMS data identification, primarily based on two strategies (Gu and Buckley, 2018). One is the expert system, which utilises decision-making strategies and logic based on the presence or absence of specific markers. These expert-system approaches often require protein sequence (.fasta) as inputs to enhance confidence in marker homology, making them more manageable for users with bioinformatics knowledge (Codlin et al., 2022; Hickinbotham et al., 2020). Another strategy is machine learning, where algorithms learn to become experts through a training process akin to human learning. However, the logic behind machine learning can be difficult to interpret since the final species assignment is determined by the votes of multiple decision trees (Baker et al., 2023; Gu and Buckley, 2018). These automated approaches are typically published as code, which can pose a learning curve for users without programming backgrounds. In the case of Baker et al. (2023) and Gu and Buckley (2018), the code was not fully published; thus, others cannot use or test these pipelines.

Given the growing popularity of ZooMS in archaeological research, with the recent establishment of many laboratories around Europe and North America, a user-friendly solution for high-throughput ZooMS identification is urgently needed. This was the second objective of this doctoral research. Here, I proposed a semi-automated ZooMS data processing pipeline. It utilises a graphics-based tool named “QuickID” to align peptide markers of samples and references, and then generate matched records. Nevertheless, the ultimate decision still depends on the analyst. The performance of this pipeline was then evaluated using published ZooMS spectra of more than 600 samples. (Section 3.2)

1.1.3 ZooMS as a zooarchaeological approach

ZooMS is a approach for taxonomically identifying bones, akin to morphological observation in traditional zooarchaeology, with the distinction being that the latter relies on morphological features still preserved on bones. A recent double-blind experiment has verified that ZooMS and morphological identification yield statistically indistinguishable taxonomic profiles for the same archaeological faunal assemblage (Morin et al., 2023). These two approaches are not mutually exclusive; rather, they complement each other. As a relatively young technique, ZooMS leverages existing zooarchaeological knowledge to narrow down assignments to a smaller number of potential species, enhancing the resolution of ZooMS identification. Concurrently, ZooMS contributes to traditional zooarchaeological methodology by addressing some challenges in morphological identification that have persisted for decades (Driver et al., 2011). It helps mitigate biases toward dominant species and rectify inaccuracies in zooarchaeological knowledge (Morin et al., 2023a; Sjögren et al., 2023). Effective integration of ZooMS and traditional zooarchaeological studies necessitates a deep understanding of both approaches. An increasing number of zooarchaeologists are embracing ZooMS, fostering dynamic communication between researchers from diverse backgrounds (e.g. Workshop: Integrating ZooMS and Zooarchaeology, Canterbury, 2023).

Integration is the concern of extensive ZooMS datasets. To date, nearly 20 Palaeolithic sites have undergone large-scale ZooMS screening (>200 specimens), predominantly in Europe, except for Denisova Cave in Siberia. The screening at these sites was primarily driven by the potential for discovering new hominin remains. However, only a handful achieved this objective (Brown et al., 2021b; Devièse et al., 2017; Hublin et al., 2020; Mylopotamitaki et al., 2024; Welker et al., 2015). Subsequent radiocarbon dating and aDNA analyses of these newfound hominin fossils have yielded breakthrough evidence regarding human dispersals and introgression events. The remaining fauna datasets resulting from ZooMS screening were unreported initially, while some were later published as test datasets for developing ZooMS methods (Gu and Buckley, 2018; Oldfield et al., 2023; Welker et al., 2017b).

In recent years, the emphasis on diverse integration approaches on ZooMS and zooarchaeological data is more common (Brown et al., 2021c; Mylopotamitaki et al., 2024; Ruebens et al., 2023, 2022; Silvestrini et al., 2022; Sinet-Mathiot et al., 2023, 2019; Torres-Iglesias et al., 2024). These integrating endeavours are marked by

creativity, tailored to the characteristics of individual ZooMS assemblages. Comparisons between taxonomic abundance in ZooMS and previous zooarchaeological studies revealed either closely matching or distinctly different patterns (Sinet-Mathiot et al., 2019). In the study by Sinet-Mathiot et al. (2023), recently excavated faunal specimens from three distinct Middle to Upper Palaeolithic transition contexts were subjected to various routine zooarchaeological methods (e.g., bone element analysis, surface readability, weathering stage, modifications, etc.), with all unidentifiable specimens analysed using ZooMS. This comprehensive integration enhances insights into hominin subsistence strategies during the transition period. However, it necessitated the exclusion of small specimens measuring less than 2 cm.

Glutamine deamidation is a common protein post-translational modification (PTM) in ancient material (Creecy et al., 2021; van Doorn et al., 2012). The isotopic distribution in MALDI-TOF-MS spectra reflects the deamidation level of a given peptide. Deamidation level serves not only as a proxy for the taphonomic environment and thermal age but also as an indicator of collagen preservation and outliers in ZooMS studies (Peters et al., 2023; van Doorn et al., 2012; Welker et al., 2016). The deamidation level revealed by ZooMS provides a quantifiable indicator of bone preservation from a molecular perspective, contributing additional detail to zooarchaeology.

As the volume of ZooMS data continues to grow, the third objective of this doctoral research was formulated. This involved the extensive ZooMS screening of Palaeolithic assemblages from Vogelherd Cave (Germany) and Jinsitai Cave (China), resulting in ZooMS datasets that were comparable to previously conducted zooarchaeological studies. The third objective focused on the integration of ZooMS data and traditional zooarchaeological results, particularly within the Palaeolithic context. It also included novel perspectives such as deamidation level. The ultimate goal was to establish a foundation for a framework that harmonises various zooarchaeological approaches and datasets. (Manuscripts B and C)

1.2 Limitations

There are two main limitations in this work. The first related to the experiment designed to assess multiple protocols. While the study compares three established ZooMS

protocols with the SP3 protocol derived from proteomics, a notable concern emerges from the adaptation of SP3 to ZooMS analysis. The SP3 protocol is distinct from routine ZooMS protocols in various aspects, from the demineralisation solvents to the purification method. As a result, the observed subpar performance can only be attributed holistically to the whole protocol, rather than evaluating more specific factors. For example, steps like reduction and alkylation in SP3 protocol are designed for broader protein extraction and become redundant in the context of collagen-only extraction. Additionally, the utilisation of EDTA as a demineralisation buffer in the SP3 protocol retains its appeal due to its compatibility with trypsin, which simplifies the ZooMS protocol (Jensen et al., 2023).

The second limitation of this study revolves around the absence of stratigraphic information and individual finding numbers for most samples. These bone fragments used to be deemed insignificant by both excavators and curators. When subsampling for ZooMS, we usually coded them from zero, which can result in codes overlapping with prior zooarchaeological studies. The absence of stratigraphic information and finding numbers imposes constraints on integrating ZooMS and previous zooarchaeology data stratigraphically. However, this scenario is a common occurrence, especially concerning unidentifiable old fauna collections, which form the majority of samples in ZooMS projects. This limitation also underscores the pressing need for improved excavation practices, comprehensive curation and nomenclature standards, ensuring meticulous documentation of the archaeological material's provenance, particularly in China.

1.3 Organization of the dissertation

This dissertation consists of five chapters. Chapter 1 presents the objectives and limitations of the study. Chapter 2 outlines the materials and methods used in this research. Chapter 3 has results in three sections: 3.1 is the evaluation of the protocol outlined in Manuscript A; 3.2 assesses the efficiency of a semi-automated identification pipeline; 3.3 presents the results of ZooMS applications in Europe and East Asia. Chapter 4 provides the discussion relating to Manuscripts B and C on the ZooMS applications and integration with traditional zooarchaeology. Finally, Chapter 5 is the conclusion and future considerations of this research.

2. Materials and methods

2.1 Materials

This doctoral research includes ZooMS analysis and publications mainly on nine archaeological sites (Figure 1). However, the majority of the samples analysed come from two old faunal collections: one originating from Vogelherd Cave in Germany and the other from Jinsitai Cave in China. Additionally, a preliminary investigation of fauna from Yumidong Cave, also in China, was carried out to assess the viability of applying ZooMS in southern subtropical locations.

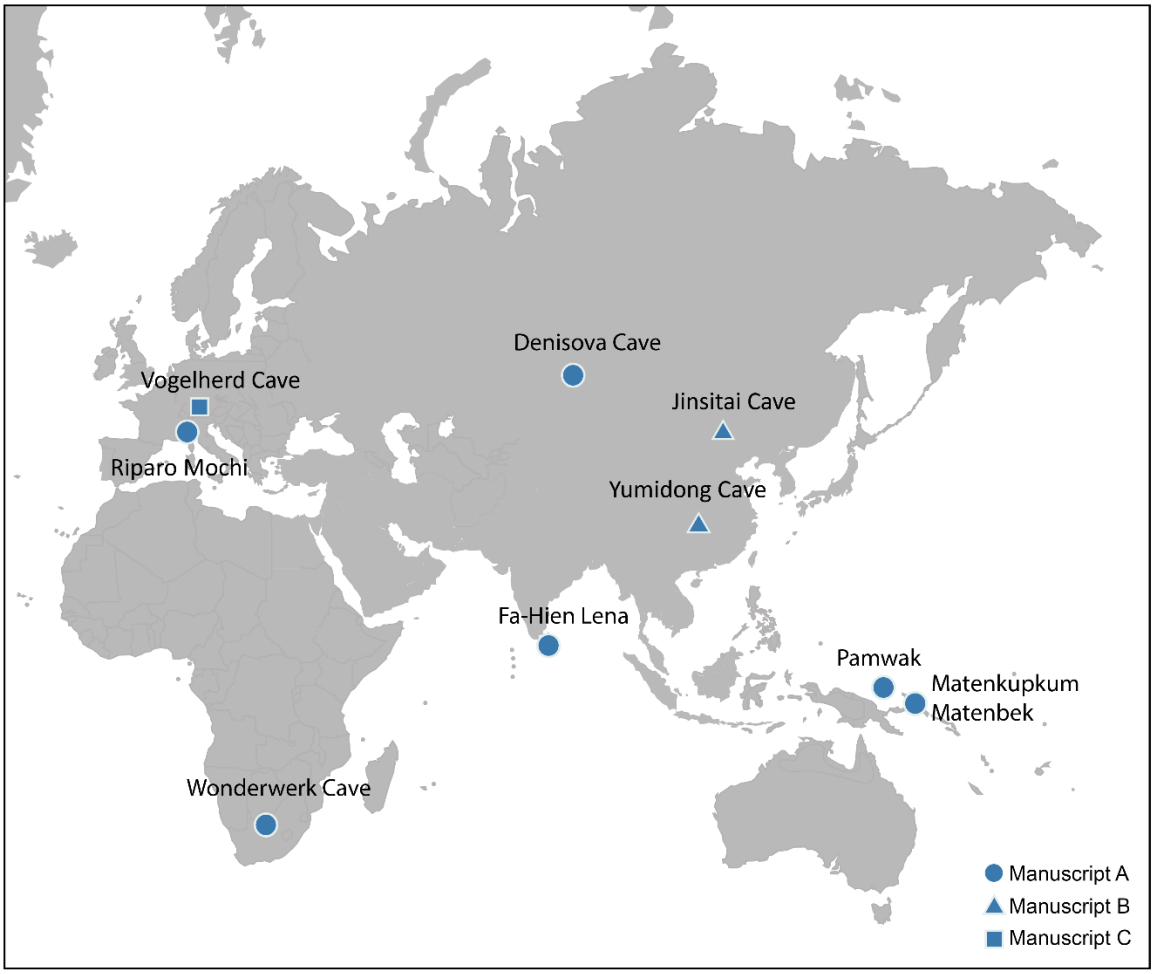


Figure 1: The location of sites studied in manuscripts (adapted from Wang et al., 2021).

2.1.1 Vogelherd Cave, Germany

Vogelherd Cave is located in the Swabian Jura region of southwestern Germany. The cave covers an area of approximately 170 m^2 and has three entrances, making it a desirable location for human habitation. The excavation in 1931 by Riek showed that the deposit was formed from the Middle Palaeolithic to Neolithic, with the two Aurignacian horizons being the richest (Riek, 1934). Besides thousands of lithic artefacts, over 18,000 fauna remains and 11 ivory figurines were recovered from the Aurignacian horizons, representing one of the largest Upper Palaeolithic faunal assemblages in Europe (Hahn, 1977; Niven, 2007). The complete figurines include an ivory mammoth with a perforation between the legs, representatives of a high art level for the time. Along with other human remains, Riek recovered a modern human skull, known as the Stetten 1 cranium, from the base of Aurignacian Horizon V (Riek, 1934). This fossil long served as evidence for assigning the figurative art to modern humans. However, direct radiocarbon dating of these human bones revealed that they were of the Neolithic period, rather than Aurignacian as initially assumed (Conard, 2009; Conard et al., 2004).

The 1931 excavation at Vogelherd was completed in twelve weeks and inevitably overlooked a large number of remains. From 2005 to 2012 and later from 2022 to 2023, Nicholas J. Conard and the team from the University of Tübingen unearthed the backdirt of the 1931 excavation, on a slope outside the cave. In the 2005 to 2012 excavation, large and visible bone remains (and other artefacts) were plotted and individually recorded. At the same time, every 5 cm-thick spit of sediments (sub-square corner) was collected in buckets. Over 32,000 buckets of sediments were then water-screened using a 2 mm sieve. The dried finds was carefully sorted: identifiable faunal bones (including microfauna), teeth/ivory fragments, burnt bones and small artefacts, such as ivory beads, were separated from highly fragmented bones. Hundreds of ivory beads were found in the backdirt, similar to the Aurignacian beads from other Swabian Jura caves (Hahn, 1988; Velliky et al., 2021; Wolf, 2015).

To recover more human remains, we designed a ZooMS screening project. Samples were selected from the water-screened bones assemblage of spits containing ivory beads, outside the cave's southwestern entrance. In this screening project, 287

unidentifiable bone fragments were sampled for ZooMS, each weighing >70 mg and mostly between 1 and 2 cm in length.

2.1.2 Jinsitai Cave, China

Jinsitai is a granite cave located at the eastern end of the mid-latitude, semi-arid Eurasian belt, 25 km from the China-Mongolia border. The cave covers an area of nearly 120 m^2. Initial excavation in 2000-2001 removed most of the 5 m thick deposit, leaving behind only the witness columns between squares. According to the published report, around 5,000 lithic artefacts, 3,000 faunal remains and three hearths were recovered in the excavations (Wang et al., 2010). Six Palaeolithic layers were identified, the upper two contained microblades and bifacially thinned points, alongside the traditional core-and-flake (small flakes) industry, which is typical in contemporary sites in northern China. This industry dominated the lower layers, with a few distinctive flakes produced using the Levallois method. In Europe and northern Eurasia, the Levallois technique is usually linked to Middle Palaeolithic Neandertals, although it is also used in some Initial Upper Paleolithic and early Upper Palaeolithic lithic assemblages. Jinsitai yielded the most typical *in situ* Levallois lithics in China, potentially indicating a cultural, and/or demic, diffusion of the technique from the west, that is distinct from the local lithic tradition (Li et al., 2018).

The initial excavation at Jinsitai lacked detailed recording, thus constraining further discussions on the lithic industry at Jinsitai. A subsequent excavation was carried out in 2012, focusing on the excavation of the remaining deposit. Limited by the digging area, the re-excavation yielded fewer finds. However, employing standard Palaeolithic excavation methods, the new campaign linked specific lithic industries and layers more confidently and provided material for radiocarbon dates. The lowest two layers with Mousterian-like lithics were dated to 47-37 cal ka BP (based on bone collagen) (Li et al., 2018).

According to the lithic comparison published in recent years (Khatsenovich et al., 2023; Rybin and Khatsenovich, 2020), the lower layer assemblage at Jinsitai is most comparable with the Middle Palaeolithic layers at Tsagaan Agui Cave in Mongolia and Denisova Cave in Siberia. Moreover, the age of the lower layers at Jinsitai is consistent with the major dispersal of modern humans to East Asia (Li et al., 2019; Yang et al.,

2024), which makes Jinsitai a key site in understanding the Palaeolithic record of the region. Given the absence of any morphologically identified human fossils at Jinsitai, we used ZooMS to screen the fauna assemblage for potential human bones in this project.

In the 2000-2001 campaign, excavators collected all bones by layers, and only those bearing morphological features were labelled. After the examination by zooarchaeologists, all unidentifiable fragments were mixed together. The entire mixed assemblage was sampled for this study, comprising 745 bone samples.

2.1.3 Yumidong Cave, China

Yumidong Cave is a large horizontal karst cave located in Three Gorges region (latitude range 29°30′–31°30′N, longitude range 106°30′–111°30′E). The Three Gorges region is a hotspot for Pleistocene age localities in China (Pei et al., 2013), with over 100 Palaeolithic sites in the region, including Longgupo Cave, which yielded crucial archaic hominin fossils (Han et al., 2017).

Yumidong Cave consists of a main chamber and a narrower branch chamber. A vertical 3-m-wide skylight provides air circulation and light, making the cave attractive to human occupation. Excavations at Yumidong began in 2011 at the area between the roof skylight and the cave entrance. Approximately 150 m² of surface area was exposed. It was dug to the depth of 5 m without reaching the bedrock. The excavation yielded thousands of lithic and fauna remains. The large limestone lithics at Yumidong belong to the cobble industry, which prevailed in southern China throughout the Palaeolithic age (Bar-Yosef et al., 2012). Multiple dating methods were employed to date the sequence, including AMS radiocarbon, U-series and ESR techniques. The Bayesian analysis of 48 dates established a geological and archaeological record spanning ca. 300 ka at Yumidong from the Middle to Late Pleistocene (Shao et al., 2022).

Unlike Vogelherd or Jinsitai, Yumidong was sampled for a ZooMS pilot study, primarily aimed at testing collagen preservation within the subtropical region where the site is located and assessing the feasibility of ZooMS. Yumidong is a recently excavated cave, and every find has stratigraphic information, including the bone fragments we analysed. Moreover, the well-established chronological framework associated with

Yumidong further heightened its appeal as a research site for a regional pilot study. From layers 2-9, we selected ~15 specimens of each layer for ZooMS, totalling 121 bone fragments.

2.2 Methods

Ancient bones, especially of potential human origin, cannot be exported from China. Hence, the FINDER project has cooperated with the School of Life Science, Jilin University, to establish a palaeoproteomics laboratory there. This is the first laboratory in China devoted to the palaeoproteomics analysis. The remaining (non-Chinese) samples were analysed in the ZooMS Laboratory of Max Planck Institute of Geoanthropology (Jena, Germany). Nitrogen and radiocarbon isotope sub-samples were sent to the Oxford Radiocarbon Accelerator Unit (University of Oxford, UK) for measurements, following published protocols (Brock et al., 2012, 2010).

Each bone was subsampled using a circular diamond saw blade. To eliminate surface contaminants, a small area of the bone was sandblasted before removing ~2 mg for nitrogen measurement and ~20 mg for ZooMS analysis.

2.2.1 ZooMS chemical protocols

In this study, we implemented four ZooMS protocols. Manuscript A is tailored to assess the effectiveness of these four protocols. Following the evaluation of these results, the AmBic protocol was applied to all 287 Vogelherd samples. Those that failed to produce high-quality spectra (n=123) underwent re-analysis using the Acid-insoluble protocol. All Chinese samples underwent analyses directly with the Acid-insoluble protocol. With the exception of the protocol evaluating phase, all other projects involved the process using 96 dropwells after the demineralisation step, a workflow that ensured a high-throughput approach. In more detail:

AmBic protocol: Each bone sample was immersed in 100 µL 50 mM ammonium bicarbonate (NH_4HCO_3) at room temperature overnight. The supernatant was discarded and an additional 100 µL of NH_4HCO_3 was added. Following incubation at 65 °C for 1 h, the supernatant was collected and digested with 0.5 µg trypsin at 37 ℃ for 18 h. The digestion ended with the addition of 1 µL 5% trifluoroacetic acid (TFA).

The digested analytes were concentrated and desalted using C18 ZipTips. The resulting elution was diluted 10 times and mixed with an equal volume of CHCA solution (10 mg/mL in 50% ACN/ 0.1% TFA (*v/v*)). 1.5 µL of the mixture was spotted on a Bruker ground steel plate prior to MALDI-TOF-MS analyses (Autoflex Speed LRF MALDI-TOF-MS, Bruker; 5800 MALDI-TOF/TOF-MS, AB Sciex).

Acid-insoluble protocol: Samples were demineralised in 500 µL 500 mM HCl for 1–5 days at 4 °C until the bone chips became spongy and stopped reacting. The supernatant was collected in clean tubes and was not used further in this protocol. The bone chips were rinsed three times with 200 µL 50 mM NH_4HCO_3 until reaching a neutral pH. The samples were then incubated at 65 °C for 1 h, in 100 µL of 50 mM NH_4HCO_3. Trypsin digestion, desalting and spotting were the same as outlined for the AmBic protocol.

Acid-soluble protocol: 500 µL of the collected HCl supernatant (see in acid-insoluble protocol) was moved to a 30 kDa ultrafilter. Centrifugation allowed the removal of the HCl and any low-weight molecules, a total volume of 750 µL 50 mM NH_4HCO_3 was used to rinse each sample 3 times. Finally, another 100 µL NH_4HCO_3 was added to recover the high-weight fraction remaining in the ultrafilter. Trypsin digestion, desalting and spotting were the same as outlined for the AmBic protocol.

SP3 protocol: Samples were demineralised in 500 µL 500 mM ethylenediaminetetraacetic acid (EDTA) for 5 days on a rotator. 100 µL of the supernatant was decanted and mixed with 200 µL 6 M guanidine HCl for protein solubilisation. Prior to binding to the magnetic beads, any proteins were reduced and alkylated for 10 min at 99 °C in 15 µL 10 mM Tris (2-carboxyethyl) phosphine (TCEP) and 15 µL 10 mM chloroacetamide (CAA). Hydrophilic and hydrophobic paramagnetic beads were used in a ratio of 1:1 (*v/v*). 20µL of 20 µg/ µL bead mix was added to each sample. Following this, 350 µL 100% ethanol (EtOH) (*v/v*) was added to reach a final concentration of 50% ethanol in the tube, proteins were allowed to bond with the beads for 5 min in a ThermoMixer. The sample-beads mixture was placed on the magnetic rack for 2 min to collect the beads and the supernatant was discarded. The residue was rinsed three times with 200 µL 80% EtOH (*v/v*) 3 times in the same way. The beads were resuspended in 75 µL of 100 mM NH_4HCO_3 and digested with 0.4 µg trypsin at 37 °C. The digestion ended by adding 10 µL of 5% TFA on the magnetic

rack and the supernatant was collected. The desalting and spotting were the same as outlined in the AmBic protocol.

2.2.2 Semi-automated data processing pipeline

To analyse the large number of mass spectra generated in this doctoral research, a semi-automated pipeline was established. It consists of three main steps: spectra peak-picking, QuickID marker alignment, and manual decision-making. Peaks in spectra were picked using flexAnalysis (V.3.4) and mMass (V.5.5.0) (Niedermeyer and Strohalm, 2012). The taxonomic identification was determined with the assistance of a tool QuickID, which aligns the sample peptides to the published ZooMS reference database. Each taxonomic assignation was then manually checked with the corresponding spectrum.

QuickID is specifically designed for ZooMS identification, based on an expert system approach, but does not require the input of protein sequences. It follows the same matching workflow as the manual identification (Brown, 2021). QuickID also features a user-friendly graphical interface. By performing binary matching with an editable marker reference library, QuickID can accelerate the most time-consuming task of ZooMS analytical workflow: the manual identification of spectra. With marker combinations output by QuickID, analysts can make quick taxonomic assignments.

QuickID requires two files as inputs: deisotoping **peak lists** (.txt), generated by spectrum analysis software such as mMass and flexAnalysis, and a **reference library** (.xlsx) collected from ZooMS reference publications. The parameter setting in QuickID (Figure 2) considers the mass tolerance and the minimum matched number (of markers). These settings play a crucial role in the accuracy of the identifications.

The default setting, illustrated in Figure 2, uses 0.3 *m/z* as the acceptable tolerance. Like many proteins, COL1 may occur post-translational modifications (PMFs), which result in mass shifts on these modified peptides. For markers with glutamine deamidation positions like P1, A, E, and G, the tolerance is adjusted to -0.3 to 1.3 *m/z*, as the PMF shift causes ca. +1 *m/z*. For markers with hydroxylation position (A, F and G), they can appear as singular peaks, or as pairs (A and A'; F and F'; G and G') due to the ca. +16 *m/z* PMF shift on some peptides. The minimum number of matched markers is set to 4 as default, meaning that any spectrum has a marker combination

containing less than four markers should be considered a failure, thus no corresponding reference records would be displayed in the output.

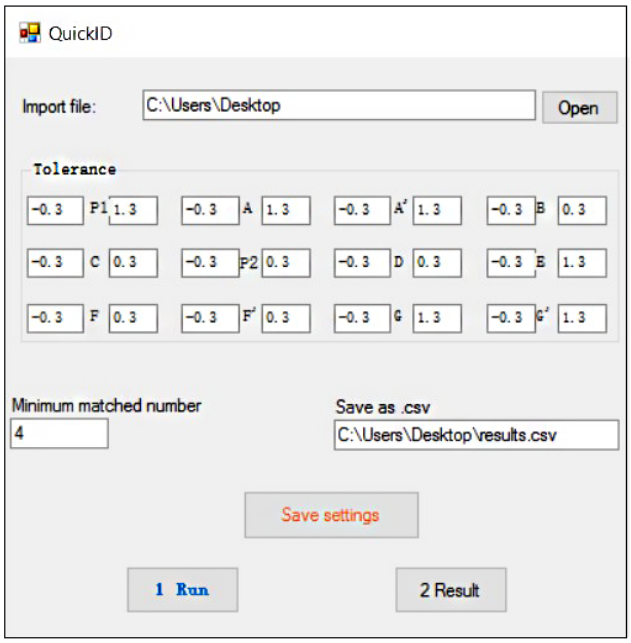


Figure 2: The QuickID interface and default settings.

Once all settings are configured, pressing "Run" initiates the matching process between peak lists and the reference library. QuickID matches the masses with reference records based on the following sequence: B, D, A, C, E, F, G, P1 and P2. Any mass in the peak list that falls within the tolerance range is recorded. The logic behind matching marker pairs is more complicated. As mentioned above, markers with hydroxide modifications can appear as pairs due to the ca. +16 *m/z* shift. However, several pair markers may cause confusion as these masses derive from different peptides, overlapping with other species, e.g.,

F: GLTGPIGPPGPAGAPGDKGE**S**GPSGPAGPTGAR (no PMF)

and

F': GLTGPIGPPGPAGAPGDKGE**A**GPSGPAGPTGAR (one Proline hydroxylation),

both sequence have a mass of 2869.4 *m/z* (Welker et al., 2016). Markers may also overlap with non-marker collagen peptides (e.g., A 1182.6 *m/z* and G 3017.5 *m/z*) (Janzen et al., 2021). To include all possible markers, QuickID retains the peak that

matches either of the pairs. When the two pair markers contradict each other, a warning is shown in the output.

Easy Label	Common Name		P1	A	A'	B	C	P2	D	E	F	F'	G	G'	Notes
03K1.txt		5	1105.6/1106.6			1453.8			2144.8		2869.1	2869.1			F and F' not match
Bear	Giant panda	Ailuropoda me	1105.6	1205.6	1221.6	1453.7	1566.8	(?)	(?)	2820.4	2853.4	2869.4	2957.5	2973.5	
Rhino	White rhinocero	Ceratotherium	1105.6	1182.6	1198.6	1453.7	1550.8	1623.8	2145.1	2820.4	2869.4	2885.4	2983.5	2999.5	
Rhino	Woolly rhinocer	Coelodonta an	1105.6	1182.6	1198.6	1453.7	1550.8	1623.8	2145.1	(?)	2869.4	2885.4	2983.5	2999.5	
Rhino	Sumatran rhinoc	Dicerorhinus s	1105.6	1182.6	1198.6	1453.7	1550.8	(?)	2145.1	(?)	2869.4	2885.4	2983.5	2999.5	
Rhino	Black rhinocero	Diceros bicorn	1105.6	1182.6	1198.6	1453.7	1550.8	(?)	2145.1	(?)	2869.4	2885.4	2983.5	2999.5	
Rhino	Javan rhinocero	Rhinoceros so	1105.6	1182.6	1198.6	1453.7	1550.8	(?)	2145.1	(?)	2869.4	2885.4	2983.5	2999.5	
Rhino	Indian rhinocer	Rhinoceros un	1105.6	1182.6	1198.6	1453.7	1550.8	(?)	2145.1	(?)	2869.4	2885.4	2983.5	2999.5	
Rhino	Sumatran rhinoc	Dicerorrhinus sumatrens		1182.6	1198.6	1453.7	1550.8		2145.1	2808.4	2869.5	2885.5	2983.4	2999.4	
Tapir	South american	Tapirus terrest	1105.6	1182.6	1198.6	1453.7	1550.8	(?)	2145.1	2792.3	2869.4	2885.4	2983.5	2999.5	

Figure 3: A screenshot of QuickID output file.

The matched masses are presented in an output file (.csv) as marker combinations. Figure 3 demonstrates the output of a spectrum 03K1. The second row displays the matching results for 03K1, where "5" indicates that the marker combination contains five markers: 1105.6/1106.6, 1453.8, 2144.8, 2869.1 and 2869.1 *m/z* (2869.1 is counted twice with a warning). Below the second row, all corresponding reference records from the library are displayed. With this information, analysts can make taxonomic assignments for the spectrum 03K1. In this case, there are four valid markers, suggesting that if the sample does not originate from southeastern Asia (which means the giant panda can be excluded), it likely belongs to Ceratomorpha, the common suborder of rhinos and tapirs. QuickID can process thousands of peak lists in seconds and export all results in one .csv file. It is convenient for subsequent operations by analysts, including filtering and ultimately identifying specific marker combinations.

2.2.3 Glutamine deamidation

Samples from Jinsitai and Yumidong were treated with the same Acid-insoluble ZooMS protocol and analysed on the same MADLI-TOF-MS instrument, which permits the evaluation of their deamidation levels. Deamidation is one of the most common protein PTMs of aged/ancient collagen (Creecy et al., 2021; van Doorn et al., 2012). Collagen peptides containing glutamine (Q) may undergo *post-mortem* deamidation, resulting in a mass shift of ~ +1 Da. Marker P1 (COL1a1 507-518, GV**Q**GPPGPAGPR), also called Cet1 in previous ZooMS publications, is a peptide shared by most mammals and contains only one Q site. It is hence identified at 1105.5 *m/z* (non-deamidated) and 1106.5 *m/z* (deamidated) in MALDI-TOF mass spectra. The

deamidation level of P1 is determined by multiple factors, including the bone's age, burial environment, sampling position, and even the protocols of collagen extraction involved (Simpson et al., 2016).

For MALDI-TOF spectra, the deamidation values are calculated by deconvoluting the contributions of the non-deamidated peptide's second isotope and the deamidated peptide's monoisotopic peak (Wilson et al., 2012). If α denotes the proportion of P1 peptide that is non-deamidated, let its intensity of peak 1105.5 *m/z* in the combined distribution be Int1, and that of 1106.5 *m/z* be Int2, then the deamidation value α is: α = Int1 / [Int2 + (1-*k*) Int1]. k is the ratio of Int2 and Int1 in an unmodified isotopic distribution, thus the *k* of peptide P1 is 0.6.

Consequently, deamidation values of 1 indicate that the peptide experienced no, or negligible, deamidation, while 0 means it is fully deamidated. The deconvolution of contributions is an indirect approach to calculating the deamidation value and is prone to distortion due to the presence of spurious peaks (e.g. baseline noise), and so values higher than 1 are also possible.

3. Results

3.1 ZooMS protocol testing (Manuscript A)

Manuscript A - Naihui Wang, Samantha Brown, Peter Ditchfield, Sandra Hebestreit, Maxim Kozilikin, Sindy Luu, Oshan Wedage, Stefano Grimaldi, Michael Chazan, Liora Kolska Horwitz, Matthew Spriggs, Glenn Summerhayes, Michael Shunkov, Kristine Korzow Richter, Katerina Douka. "Testing the efficacy and comparability of ZooMS protocols on archaeological bone." *Journal of Proteomics* 233 (2021): 104078.

Several chemical protocols have been reported since the invention of ZooMS. In Manuscript A, we evaluated three established ZooMS protocols on bones of different origins and collagen preservation. SP3, a new protocol, previously used in DNA and shotgun proteomics fields, was also tested on its ZooMS performance.

The bone samples examined in this project come from seven archaeological sites around the world, spanning from ca. 2,000 years to as old as 500,000 years ago. Some of the analysed sites are located in tropical and savanna environments, where collagen preservation is challenging. The amount of nitrogen in a bone sample, thought to be derived mainly from remaining collagen, was measured to predict the collagen quantity and is reported here as percent nitrogen (%N). The %N of bones from Wonderwerk Cave in South Africa was the highest in the study (mean = 2.07). It was unexpected since these bones come from the Middle Pleistocene context. The %N values from Denisova Cave in Siberia and Fa-Hien Lena in Sri Lanka were 1.30 and 0.56, respectively. The other four sites from Italy and Papua New Guinea revealed %N values close to 0.

The mildest and fastest collagen extraction protocol for ZooMS utilises AmBic (NH_4HCO_3) and avoids demineralising the bone matrix. In this project, AmBic protocol was applied to 400 samples, of which only 181 samples (all from Denisova Cave) returned positive results. Subsequently, using protocols with HCl to facilitate collagen extraction, we achieved 14 additional ZooMS identifications: 11 from Fa-Hien Lena and 3 from Denisova Cave. The SP3 protocol (EDTA demineralised) failed to improve

the collagen extraction success and showed the worst performance among all protocols.

A clear pattern was shown when we projected 400 samples' %N and their total peak numbers in spectra: the higher the %N, the greater the likelihood of identifying this sample using ZooMS (Figure 4 of Wang et al., 2021, Manuscripts A, page 68). Hence, if a bone contains >0.26 %N, it is likely to be identifiable using ZooMS. The ZooMS threshold is lower than that used in the radiocarbon community, which usually requires >0.76 %N for radiocarbon dating. It means if a sample fails radiocarbon dating, it is still possible to obtain a ZooMS identification, particularly when the acid demineralising protocol is employed.

The exceptions to the correlation between %N and ZooMS identificability were from Wonderwerk Cave: despite the relatively high %N values in the bones, ZooMS was unsuccessful. We suggested that the high %N values in Wonderwerk samples may come from the contamination of the bone matrix by bats/bird guano, which enriches nitrogen.

3.2 Assessment of the new pipeline

3.2.1 Generation of the QuickID-based dataset

To evaluate the new semi-automated pipeline, I downloaded a subset of published ZooMS data, manually identified by Brown et al. (2021c), as a test dataset. The dataset includes raw spectra (using AmBic protocol, in triplicate) and identifications for 651 bone samples.

The default parameters in flexAnalysis and QuickID were used to generate the new dataset for comparison (*s/n*=2 for flexAnalysis pick-picking, tolerance -0.3 to +0.3 *m/z* for QuickID matching). QuickID listed identifications based on marker combinations in the output files (link see Appendix 5). Each sample got three independent identifications. In cases where two of the three repeats failed, the sample was determined to fail. Of the 651 samples, 630 samples had three identical or inclusive assignments; these were merged into one. However, 21 samples had conflicting identifications in the three repeats. Upon manual spectra inspection, I found 20 of the

conflicts were due to false positive peaks picked by flexAnalysis. Then, I corrected these identifications manually. A case of contamination was also identified: DC8008, exhibiting two incompatible B markers (1427.5 and 1453.8 *m/z*).

3.2.2 Comparison with the manual dataset

By comparing, 508 of the 651 samples showed the same identifications in manual and QuickID-based datasets, including 354 identifiable and 154 failed samples (Table S1, link see Appendix 5). The remaining 143 samples, had different identifications, the difference pairs fell in three patterns: (i) conflicting identifications in two datasets (n=1), (ii) inclusive identifications in two datasets (n=56), and (iii) sample failed in one dataset but was identified in the other (n=86).

I examined the spectra of these 143 samples to understand the inconsistencies and thus evaluate QuickID's performance. Sixteen samples had problematic manual identifications (including the single conflicting case). The QuickID-based pipeline was responsible for 14 non-compatible results. These inconsistencies arose from the limited mass tolerance (n=13) and false positive peaks (n=1). The remaining differences came from different standards for failed/identified samples (n=70), and the variability in few broad assignations (Ovis/Capra/Cervidae/Gazella/Saiga, n=39) during manual identification. Additionally, four samples gave problematic assignments in both datasets. Overall, the QuickID-based pipeline is more reproducible than the manual one in ZooMS identifications.

Figure 4 exemplifies the factors affecting the accuracy of an identification. One challenge comes from the presence of low-intensity peaks (in the context of *s/n*=2). Some peaks are picked into the peak lists but are not truly visible in the spectra (false positive, Figure 4, a); meanwhile, some low-intensity peaks cannot pass the *s/n* threshold during the peak-picking process and not picked (false negative, Figure 4, b and c). The former scenario is more damaging than the latter in ZooMS identification. False positive/negative peaks occur randomly and thus tend to cause conflicting identifications in the triplicate runs. Hence, false peaks are usually noticeable during merging the triplicates in the QuickID-based pipeline.

Another challenge stems from the mass shifts. The mass tolerance setting in QuickID is crucial for correct identifications. The default setting is -0.3 to +0.3 *m/z*. In some

cases, the observed mass may fall out of the tolerance range and not be picked. The omissions in peak-picking also result from monoisotopic mislabelling or insufficient calibration (Figure 4, d and e), resulting in identifications of lower taxonomic resolution. Manual identification is less affected by mass shifting/mislabelling, compared with QuickID.

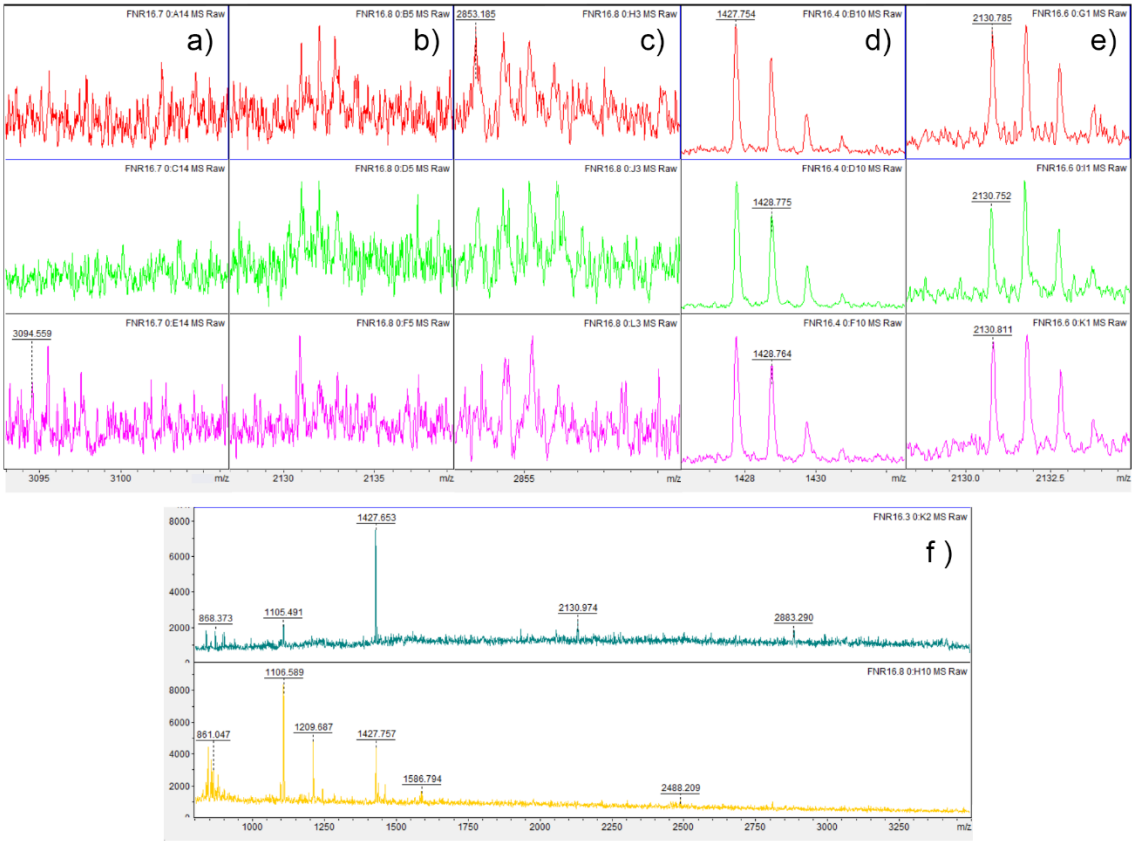


Figure 4: Mass spectra of (a) DC7535 triplicate: The 3094 *m/z* in the bottom spectrum is a false positive peak; (b) DC7635 triplicate: The 2131 *m/z* is not picked. (c) DC7680 triplicate: The 2853 *m/z* is only picked in one spectrum; (d) DC8208 triplicate: The monoisotopic peaks at 1427 *m/z* are mislabelled in two spectra; (e) DC8400 triplicate: Marker D in two spectra shifts over 0.3 *m/z* from the 2131.1 *m/z* in the reference; (f) DC8102 (turquoise) and DC7695 (yellow): DC8102 has four markers. It was assigned as failed in the manual dataset but was identifiable in the QuickID output. DC7695 has three markers, published as manually identifiable but failed in QuickID.

Most of the inconsistent identifications are the result of different standards used in the ZooMS identification (109/143). For example, the failed/identified standards are not the same in QuickID and manual identification: both DC8102 (Figure 4, f, turquoise) and DC7695 (Figure 4, f, yellow) have poorly preserved collagen and yielded only a few markers. In the manual dataset, DC8102 failed, while DC7695 was identified as Bison/Yak based on the presence of the unique 1208 *m/z* marker. However, QuickID

requires at least four markers for an assignation, regardless of the taxon. Consequently, in the QuickID-based dataset, DC8102, with four markers, received a broader assignment of Ovis/Capra/Cervidae/Gazella/Saiga, while DC7695, with only three markers, was classified as failed manually. This kind of discrepancy arises mostly in poorly preserved samples that are on the limit of being failed or identified. In some cases, re-running using acid-based protocols can significantly improve the spectra quality.

In summary, the abovementioned problems do not significantly impact the validity of the QuickID-based semi-automated pipeline I proposed in this research. Although occasional picking of false peaks is inevitable for any spectra processing algorithm, most false peaks that affect the identification can be eradicated after manually merging all triplicates, as an advantage of the semi-automated pipeline. Suggestions for improving the pipeline performance include adjusting QuickID's default mass tolerance and flexAnalysis' signal-to-noise value, making these parameters more batch-specific. For the issues related to low-abundance peaks and failed/identified standards, the inconsistency between QuickID-based and manual identifications is not greater than that between analysts and other pipelines (Mylopotamitaki et al., 2024). Moreover, re-running the bone chips that failed in AmBic extraction with acid demineralisation is also an option to improve the quality of the spectra and the accuracy of identification.

3.3 ZooMS results

3.3.1 Vogelherd Cave

Collagen preservation at Vogelherd is known to be relatively good, as indicated by numerous successful radiocarbon dates. Based on the results of the protocol testing presented in Section 3.1, we chose the AmBic protocol for Vogelherd bone fragments in the first round of ZooMS. Nearly 40% of spectra failed to provide robust identifications. In a second round, we used the Acid-insoluble protocol; this improved the identifiable rate to 84% (241/287), also resulting in the identification of three hominin fossils (VH 057, VH 111 and VH 197, photographs see Figure 5). The three were tiny bones weighing 170-407 mg before ZooMS sampling. Direct radiocarbon dating and ancient DNA analysis suggested they belonged to at least two individuals;

the oldest one (VH 057) came from a Magdalenian individual (~16 cal ka BP), and VH 111 fell in the Neolithic period (~5 cal ka BP). VH 197 failed radiocarbon dating but genetically clustered with the European Neolithic farmer-related population (Posth and He, personal communications, 2022; Bloss et al., in prep.).

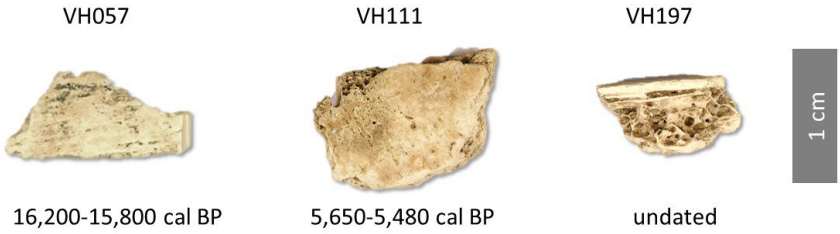


Figure 5: Photographs of the hominin bones from Vogelherd water-sieved assemblage, identified using ZooMS.

In addition to the three human remains, 199 bone fragments were identified to at least the genus level, and 42 had more generic taxonomic identifications. Although ivory fragments were excluded from the analyses, we found 32.2% of Elephantidae bone remains in the ZooMS assemblage, following 18.4% of horses and 11.4% of reindeer. The ZooMS water-screened assemblage also revealed a significantly higher number of hares (9.4%). Since the number of ZooMS-identified fragments is much smaller than the plotted finds of previous zooarchaeological studies, no new species became identified by ZooMS.

3.3.2 Jinsitai and Yumidong caves

To obtain the best quality spectra, all 866 samples from Jinsitai and Yumidong caves were analysed using the Acid-insoluble protocol. Despite the antiquity and location of the studied sites in the subtropical and temperate zones of East Asia, the ZooMS-based identifiable rates were unexpectedly high. Of the 745 bones analysed from Jinsitai Cave, 90% had sufficient collagen for an order or genus level taxonomic assignment. The Jinsitai fauna is attributed to the *Mammuthus–Coelodonta* faunal complex. Although no mammoth was identified, ZooMS resulted in 121 woolly rhinos (18.0%), after horses and bison (28.3% and 18.6%, respectively).

The ZooMS results of Jinisitai highlighted the presence of two previously unidentified species at the site. ZooMS spectra of 31 bone fragments matched with the reference of *Camelus ferus* (wild bactrian camel). Camel remains comprised 4.6% of the Jinsitai

fauna assemblage, and all specimens were highly fragmented. Radiocarbon dating of five camel bones indicated that camel presence was linked to colder periods of Marine Isotope Stages 3 and 2 (37-20 cal ka BP).

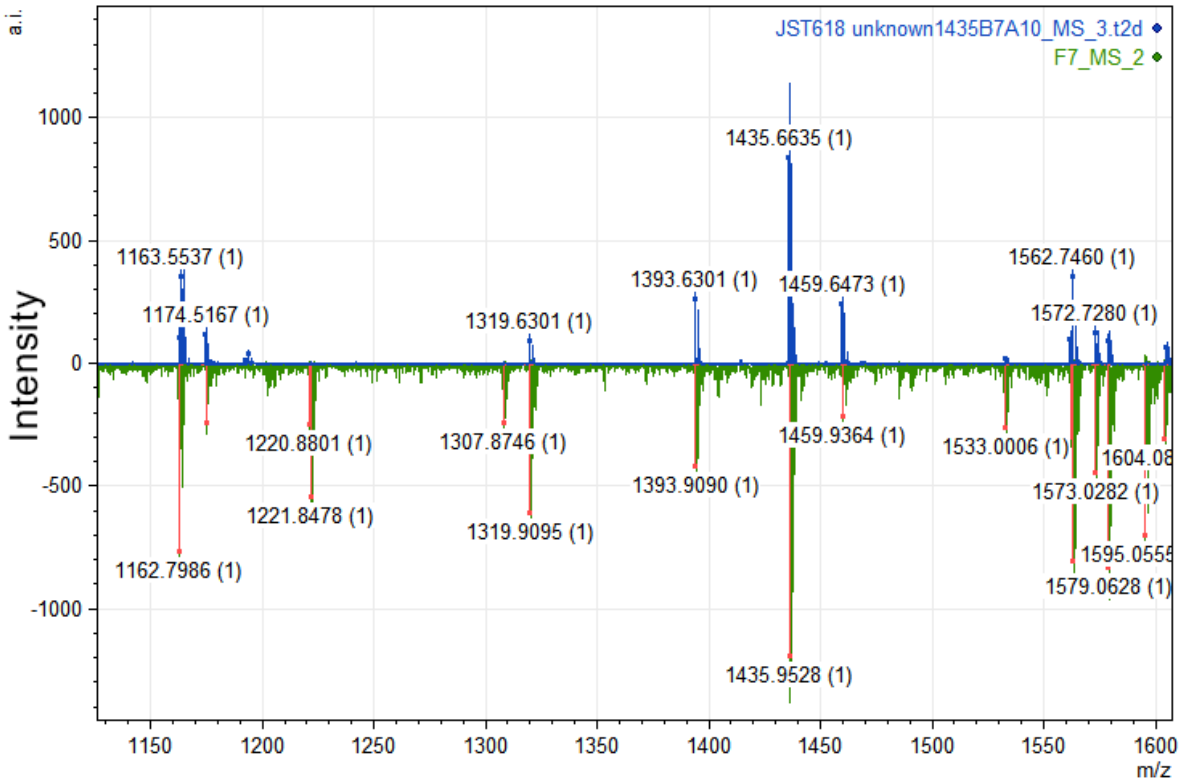


Figure 6: The spectrum above (blue) is JST 618 from Jinsitai Cave, and the one below (green) is a modern ostrich, zoomed in at 1100-1600 *m/z*.

Another identified species was not only new for Jinsitai, but also for ZooMS. Two specimens (JST 459 and 618) yielded high-quality spectra but failed to match any known species in the reference. In Manuscript B, we tentatively assigned the two samples as unknown. A mass peak at 1162.5 *m/z* (or 1163.5 by deamidation, Figure 6, blue) in the unknown spectra differs from the P1 marker shared in most mammals; instead, it belongs to the P1 of some birds and marsupial taxa (Eda et al., 2020; Peters et al., 2021). After comparing the Late Pleistocene fauna of other Chinese and Mongolian localities near Jinsitai, I considered the giant ostrich a potential candidate. To test the hypothesis, B marker (COL1a2) of several genera that are phylogenetically close to ostrich, as well as human and tapir, were aligned in Geneious (Geneious version 2023.2.1, Figure 7). Only the ostrich has a unique mutation at position 1869 of COL1a2, theoretically resulting in a diagnostic marker at 1435.6 *m/z*, as observed in the Jinsitai unknown spectra. The ostrich assignment was then confirmed by the

MALDI-TOF spectra of a modern ostrich bone (Figure 6, green). Theoretical tryptic peptides of COL1a2, including 1578.7, 2149.1 and 2927.4 *m/z* (C, D, G markers), were shared in those bird species and appeared in both Jinsitai's and modern ostrich spectra. As a result, the two unknown samples can be confirmed as ostrich, a locally extinct taxon.

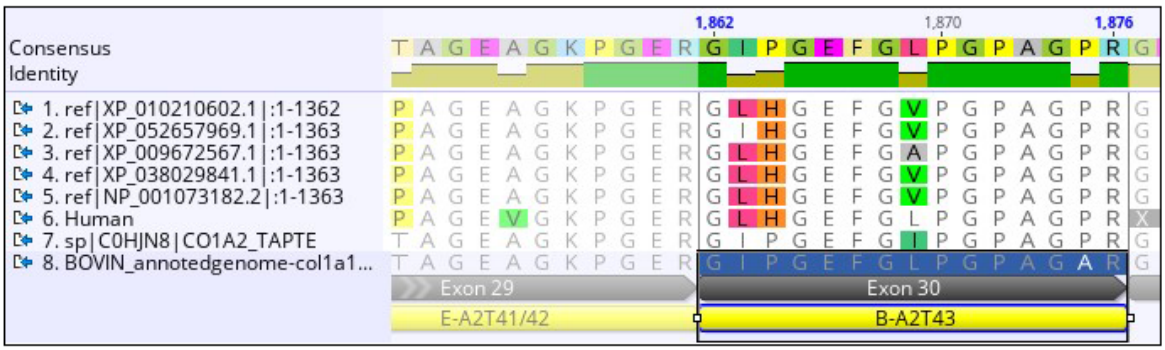


Figure 7: B marker (col1a2) alignment in Geneious, position 1862-1876. 1. *Tinamus guttatus* (tinamou, XP_010210602.1), 2. *Harpia harpyja* (harpy eagle, XP_052657969.1), 3. *Struthio camelus australis* (African ostrich, XP_009672567.1), 4. *Anas platyrhynchos* (duck, XP_038029841.1), 5. *Gallus gallus* (chicken, NP_001073182.2), 6. *Homo sapiens* (human, P08123.5), 7. *Tapirus terrestris* (tapir, C0HJN8.1)

The ZooMS identifiable rate in the pilot study on 121 Yumidong specimens was also high, reaching 83%. Yumidong fauna belongs to the *Stegodon–Ailuropoda* faunal complex, a fauna complex widely distributed from East Asia to mainland Southeast Asia during the Late Pleistocene. Elephantidae bones composed 47.0% of this ZooMS pilot assemblage, followed by 31.0% bovids. The ZooMS results of Yumidong revealed a reduced diversity of taxa, compared to the previous zooarchaeological examination, likely due to the small testing scale at this site.

The glutamine deamidation patterns of Jinsitai and Yumidong caves were noteworthy. In general, bones from Jinsitai, the younger site in northern China, were less deamidated than those from Yumidong: the median deamidation value for Jinsitai was 0.62, whereas for Yumidong the value was significantly lower at approximately 0.15. Deamidation values of each site were neither normally distributed as a whole nor distributed at distinct ranges. The deamidation values of Yumidong showed a correlation with the deposit depth: most of the deamidation variation occurred in the two upper layers, while deamidation levels in layers 4 to 9 were close to 0, meaning that the collagen was fully deamidated.

4. Discussion

4.1 Integrating ZooMS and zooarchaeology (Manuscript C)

Manuscript C - Naihui Wang, Nicholas J. Conard, Katerina Douka. “Integrating morphological and ZooMS-based approaches to zooarchaeology at Vogelherd Cave in Southwestern Germany”, submitted to the *PaleoAnthropology*. (in review)

Zooarchaeology is a fundamental branch of archaeology that has long made significant contributions to major archaeological questions. Yet, the morphology-based analysis of animal bones, or the traditional zooarchaeology approach, relies on the presence of diagnostic features on the bone. When these are lacking, alternative approaches are used to obtain taxonomic determination of bones. ZooMS has emerged in the past decade as, by far, the most accessible, cost effective approach for taxonomic identifying bones based on their collagen characteristics.

Integration of traditional zooarchaeological and ZooMS data, however, is still in its infancy. The zooarchaeology research of Vogelherd fauna has a long history, and our ZooMS data of water-screened fauna assemblage can represent another exploration of the integration.

Since the first excavation at Vogelherd in 1931, three zooarchaeological studies have improved the understanding of the fauna, ultimately informing us of the subsistence strategies of the occupants at Vogelherd and the palaeoecological framework around the site. The first fauna analysis of the 1931 collection was undertaken by palaeontologist Ulrich Lehmann in the 1950s (Lehmann, 1954). Lehmann examined 921 intact bones from the 1931 collection, and this assemblage had well-recorded archaeological contexts and retained morphological features. Through Lehmann’s morphometric measurement, some bones even reached the subspecies level identification. Elephantidae were absent in Lehmann’s assemblage as another palaeontologist was responsible for that. Lehmann’s work had a paleontological focus in the main, although he also used the results to reconstruct the palaeoclimate and human-ecological relationship, making this study closer to what was later defined as zooarchaeology.

It was not until the early 2000s that Laura Niven undertook a comprehensive zooarchaeological study of the 1931 collection (Niven, 2001, 2006, 2007). At that time, the methodology of zooarchaeology had been largely established. Niven's study incorporated 14,181 bones, the whole faunal assemblage of the original excavations. Of these, 7,730 specimens could be determined taxonomically to species to family levels. Niven applied multiple zooarchaeological methods in her study, including the inferences on accumulators, standard quantification index, modification trace, age and sex profiles, and these results were used in Niven's reconstruction of human behaviour at the site.

In 2014, after the first round of backdirt excavation at Vogelherd (Conard et al., 2012), Ulf Boger and colleagues analysed the newly piece-plotted fauna remains (Boger et al., 2014). Unlike the 1931 collection, the bones of the backdirt assemblage (n=2,342) were generally smaller in size. The initial attempt to assign these bones to their original context based on bone preservation indexes failed. Hence, the authors examined the backdirt fauna as a whole. Nearly 84% of specimens received a genus or higher level of identification, higher than that of most Palaeolithic caves. Their study also added a few new taxa to the known fauna list for Vogelherd, such as roe deer, marten, polecat, and hedgehog. The comparison of Boger et al.'s and Niven's datasets identified a higher abundance of small animals in the backdirt assemblage, with some anthropogenic modifications recognised on hare remains. Boger et al.'s analyses of the backdirt fauna complemented the previous zooarchaeological studies at the site by highlighting the potential role of low-ranked small game.

In this doctoral research, ZooMS analyses on small bones from water-sieved sediments add to the zooarchaeological body of knowledge at Vogelherd Cave. The evolution of assemblages examined in zooarchaeology unveils a trajectory from exclusively complete bones to increasingly diminutive, undiagnostic fragments. Notably, the shift in using taxonomic abundance indices at Vogelherd reflects a transition from the derived Minimum Number of Individuals (MNI, listed by Lehmann and Niven) towards the observational Number of Identified Specimens (NISP, listed by Niven and Boger et al.), in line with broader trends in zooarchaeology (Lyman, 2018). This transition is particularly advantageous for ZooMS, as its tallies align seamlessly with NISP counts.

One must exercise caution regarding the bias when comparing different assemblages. Such bias is particularly evident in the 1931 excavation, which overlooked small bone fragments (Lehmann, 1954), and the over-representation of easily identifiable specimens in the backdirt's piece-plotted fauna (Boger et al., 2014). Despite these biases, we found noteworthy similarities across the assemblages, with mammoth remains consistently contributing the highest NISP numbers, followed by dominant prey horses and reindeer. Human exploitation of horses surpasses that of reindeer, illustrating a trend supported by both NISP and MNI indices in all assemblages. The inconsistency in abundance ranking pertains to small-size taxa that are relatively low in abundance. The comparative analysis of the ZooMS assemblage with unidentified body size groups and morphologically identified assemblages exposes a greater representation of small animals in ZooMS dataset (Figures 3 and 4 of Wang et al., in prep, Manuscript C, pages 117 and 118). This observation again underscores the prevalence of high fragmentation in small game, such as hares, reinforcing a more substantial role for small game in the subsistence strategies of human occupants, after Boger et al. (2014).

The integration of morphological and ZooMS approaches necessitates a nuanced understanding of both methodologies. Morphological taxonomic results inherently encompass elemental and morphometric information, offering a comprehensive perspective. Conversely, ZooMS assemblages typically lack diagnostic characteristics. While ZooMS significantly amplifies the number of specimens with taxonomic identification, relying solely on ZooMS results can still be inadequate for explaining human behaviour. For example, without morphological observations on the high level of ivory fragmentation and cut marks on horse and reindeer bones, ZooMS results can lead to an erroneous conclusion that occupants primarily exploited mammoths at Vogelherd. Moreover, the efficiency of the morphological and ZooMS identification varies across mammalian families, and their identification advantages do not overlap uniformly. Certain bovids, for instance, are notoriously difficult to identify using ZooMS, while they have distinct differences on their horns.

ZooMS can improve morphological identification as well, given its ability to distinguish sheep/goat, bison/buffalo and donkey/horse (Buckley et al., 2010; Paladugu et al., 2023). Emerging trends include the use of specimens with specific anatomical elements or modifications in ZooMS analysis, reflecting a growing appreciation on the

use of both morphological and biomolecular approaches in the same assemblage (Pothier Bouchard et al., 2020; Mylopotamitaki et al., 2024; Ruebens et al., 2023; Torres-Iglesias et al., 2024). Furthermore, the deamidation level in conjunction with ZooMS identification, offers an additional quantifiable index for bone preservation. When combined with bone weight (or length) measurements, ZooMS approach enables powerful insights into bone fragmentation, spatial distribution, and overall preservation.

The re-excavation of backdirt extends beyond Vogelherd, and includes other archaeological sites in Europe, such as Feldhofer Cave (Schmitz et al., 2002), Svedův Stůl Cave (Nejman et al., 2020; Wright et al., 2021), Spy Cave (Pirson et al., 2012) and the Ànimes Caves (Morales et al., 2022), as well as the currently investigated backdirt of Jinsitai Cave. These are pivotal Palaeolithic sites, initially investigated decades ago with inappropriate methods in place for the retrieval of all archaeological remains. Nowadays, the use of wet/dry sieving is integral in the re-excavation of backdirt sediments. These efforts not only yield remarkable new finds, such as human fossils and artefacts, but also unearth a substantial volume of previously overlooked fauna. Our research on Vogelherd suggested that the extensive application of ZooMS is a suitable technique for taxonomically identifying small bones sieved from sediments. Combined with zooarchaeology, ZooMS has the potential to significantly enhance and paint the full picture of animal presence and exploitation at archaeological sites.

4.2 ZooMS applications in East Asia (Manuscript B)

Manuscript B. Naihui Wang, Xu Yang, Zhuowei Tang, Cunding He, Xin Hu, Yinqiu Cui, Katerina Douka. Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China. *Proceedings of the Royal Society B* 290, no. 2009 (2023): 20231129.

While ZooMS has been mainly applied to western Eurasian assemblages, its potential for other parts of the world has remained limited. At the onset of this doctoral work, almost no application of ZooMS to material from East Asia had been performed, and this was one of the main objectives of this research under the framework of FINDER Project.

The time and routes of modern human dispersal in East Asia have long been debated. Fossil evidence of early modern humans, dating from 200 to 60 ka, has been documented at various locations beyond Africa (Groucutt et al., 2018; Harvati et al., 2019; Liu et al., 2015; Shackelford et al., 2018; Westaway et al., 2017). Genetic studies on present-day human populations indicate that earlier dispersals did not contribute detectable ancestry to people living after the major global human expansion, which occurred after 60-50 ka (Bergström et al., 2021; Posth et al., 2016). In East Asia, several fossils have been identified as potentially belonging to early modern humans (Bae et al., 2014; Bae et al., 2017; Liu et al., 2015; Shen et al., 2013). Alongside early modern humans, archaic hominins have also been identified in China, including the Baishiya Denisovan, which inhabited the Tibetan Plateau from approximately 160 to 60 ka (Chen et al., 2019; Zhang et al., 2020), and the Harbin archaic hominin, dated to at least 138 ka in Northeast China (Ni et al., 2021). Furthermore, hominin fossils displaying mosaic features of both modern and archaic characteristics from southern China have been reported, with contexts dating to the end of Late Pleistocene, e.g. Liao et al. (2019). All of these finds, despite sometimes controversial, have contributed to the enigma surrounding late human evolution in East Asia.

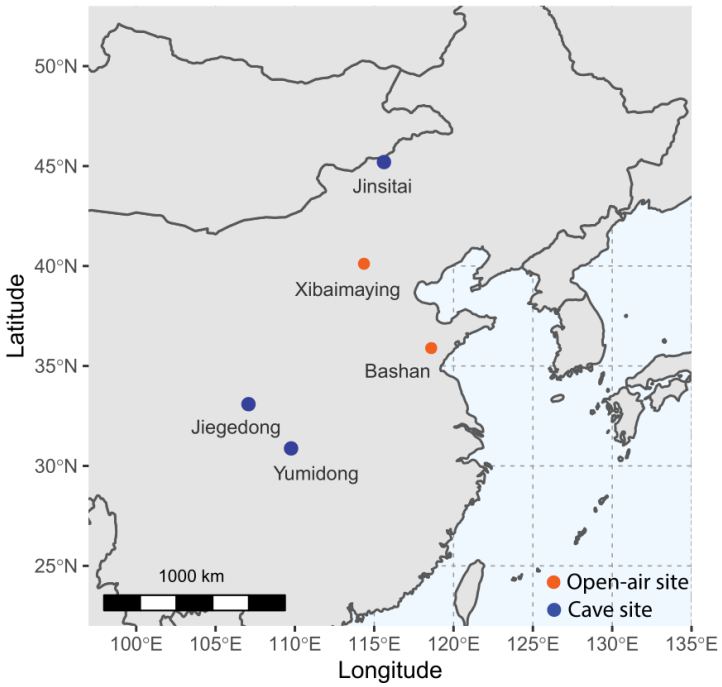


Figure 8: Sampled locations for the ZooMS feasibility studies in China (Basemap by R package rnaturalearth)

To assess the applicability of ZooMS for discovering potential new human remains in East Asia, this doctoral research involved five Palaeolithic sites in China, encompassing a large-scale screening project at Jinsitai Cave and four pilot studies (site locations shown in Figure 8). Among these sites, only Jinsitai and Yumidong have published archaeological reports and zooarchaeological studies. Remarkably, three cave sites exhibited unexpectedly high ZooMS identifiable rates, from north to south: Jinsitai on the Mongolian Plateau (90%), Jiegedong in the Qinling Mountains (88%), Yumidong in the Three Gorges (83%).

Particularly worth mentioning is the preservation condition of the two limestone caves in southern China, Jiegedong and Yumidong, both consist of horizons predating 100 ka and spanning until the end of the Pleistocene. The limestone caves prove favourable to the long-term preservation of collagen, and Jiegedong also displays an increasing identification rate from older to younger horizons (see Appendix 4).

Open-air sites are more subjected to environmental fluctuations than cave sites, leading to faster degradation of bone collagen due to frequent wet-dry/ freeze-thaw cycles. The younger open-air site Xibaimaying (ca. 50-30 ka) in the Nihewan Basin showed >60% identification success rate (n=39), while the older Bashan open-air site (ca. 100-60 ka) yielded no ZooMS results from all 63 specimens (Appendix 4). Despite the limitations of open-air sites, these pilot studies demonstrate ZooMS to be an extremely promising method for sites in East Asia.

The extensive ZooMS screening at Jinsitai involved the examination of the entire unidentified fauna collection. This work added two new taxa to the known fauna list for the site. Notably, the screening identified 31 camel specimens, constituting ~5% of the Jinsitai unidentified fauna assemblage. Their spectra matched the ZooMS reference of the extant double-humped Bactrian camel, as distinct from the single-humped dromedary camel. While uncertainties persist regarding the relationship between *Camelus knoblochi* and Bactrian camels, *Camelus knoblochi* is more likely to have been present at Jinsitai. This extinct "giant" camel species inhabited northern China for millennia and probably became extinct around 20 ka BP (Tang et al., 2003; Titov, 2008). Remains of this species have been identified at Dabusu, a palaeontological locality 500 km east of Jinsitai (Tang et al., 2003), and in the Palaeolithic sites Wulanmulun and Salawusu in Inner Mongolia (Dong et al., 2014). Until 2022, *Camelus*

knoblochi was reported in Mongolia for the first time from Tsagaan Agui Cave (Klementiev et al., 2022).

Traditionally, camels are not considered a targeted species for Palaeolithic hunters due to the rarity of camel remains in anthropogenic deposits. However, the highly fragmentary nature of camel bones at Jinsitai suggests human exploitation of camels, possibly involving selective transfer of specific body parts to the cave. Additionally, five dated camel specimens represent at least four individuals (37-20 cal ka BP), indicating that the camel's presence at Jinsitai was not a one-off encounter. Through the application of ZooMS, this research unveils to date the most numerous archaeological camel remains in the Palaeolithic context, spanning a remarkable 17 millennia. As hunter-gatherer populations expanded across North and Central Asia, they may have encountered and targeted camels, among other megafauna. Camels, which later became indispensable on the Silk Road, were also not ignorable in the Palaeolithic period.

Another newly identified taxon at Jinsitai is the ostrich, a non-flying giant Aves species. Ostrich eggs, characterised by their large size and thick, durable shells, were utilised as a valuable food resource and material by hominins in the Pleistocene. The importance of ostrich eggshells in archaeology lies in the beads made from eggshells, which serve as portable personal ornaments that may signify intellectual or social interactions among humans - an aspect considered a part of modern human behavioural evolution. Ostrich eggshell beads are predominantly found in Africa, India, Siberia, and East Asia from contexts within 50 ka, the technique and chronology of these artefacts have been extensively studied (Bednarik, 2011; Derevianko and Rybin, 2003; Miller and Wang, 2022; Song et al., 2022; Wei et al., 2017). The oldest sequenced ancient peptides and DNA have also been successfully extracted from eggshells (Demarchi et al., 2016; Jain et al., 2017).

East Asian Ostrich (*Struthio anderssoni*), was once widely distributed in East Asia but became extinct before the Holocene (Kurochkin et al., 2010). Bone remains of this taxon have been reported from a few Upper Palaeolithic caves and open-air sites in Mongolia and North China, including Tolbor (Gladyshev et al., 2010), Dörölj-1 (Gladyshev et al., 2012), Tsagaan Agui (Martynovich, 2002), Salawusu (Qi, 1975), Wulanmulun, Shuidonggou and Xujiayao (Zhang, personal communications, 2023).

Thus, the identification of two ostrich bones at Jinsitai is not unexpected, given its location in the Mongolian Plateau, a region conducive to ostrich habitation. Since only two out of 745 fauna remains are ostrich, this taxon might not be one of the megafauna prey for occupants at Jinsitai, and the potential utilisation of eggs is still ambiguous. The ongoing re-excavation of backdirt at the site, coupled with the sediment sieving process, holds promise for the discovery of additional ostrich bones or eggshell artefacts, should they exist.

Glutamine deamidation analyses in this doctoral research can help in the backdirt reconstruction of Jinsitai. The rate of deamidation, influenced by protein structure, burial temperature and pH, etc. (van Doorn et al., 2012), renders it a valuable proxy for collagen preservation, indirectly linked to age. Specific deamidation studies on large datasets from various sites have yielded variable results as to the utility of this proxy. For instance, for the Châtelperronian fauna assemblage of Quinçay (n=457), deamidation values exhibited greater spatial variation than chronological resolution (Welker et al., 2017b). In contrast, a study on the Denisova fauna assemblage spanning 250 ka showed negligible chronological variation (n=2459), possibly due to the extremely favourable environment for collagen preservation in the cave (Brown et al., 2021a). A similar pattern emerged in a small dataset of the Late Pleistocene deposits at Pin Hole, where deamidation appeared to reflect more depositional conditions than chronological factors (Buckley et al., 2017).

At Jinsitai and Yumidong, for the first time, deamidation seems to reflect more chronological factors than depositional conditions. The zooarchaeological study at Jinsitai has revealed a changing taxa composition over its occupation, with woolly rhinos predominant mainly during the early Upper Palaeolithic period, red deer and marmots appearing towards the end of Upper Palaeolithic, and pigs only found in the uppermost Neolithic horizon. Correspondingly, mean deamidation values of taxa groups mirrored this narrative (Figure 2 of Wang et al., 2023, Manuscript B, page 83), with Rhinocerotidae exhibiting the lowest mean value (indicating more deamidation), while Suidae and Rodentia (marmots) having the highest mean values (indicating lesser deamidation). Similarly, deamidation data from Yumidong unveiled distinct stages of *post-mortem* protein modification, with ongoing deamidation in the upper layers and complete deamidation in the lower layers. For indicating relative age or collagen preservation, the mean deamidation value of a group of samples is

statistically more reliable than isolated values, considering potential high variation within a group. Deamidation is not an absolute age indicator for individual samples, as demonstrated by the lack of consistent correlation between radiocarbon dates and deamidation values in five directly dated camel bones.

For Jinsitai, deamidation can continue to provide valuable insights into the reconstruction of the origin of the backdirt. The cave was mainly excavated in the early periods when archaeological fieldwork standards had not yet been appreciated. The ongoing re-excavation of backdirt will unearth numerous unstratified bone remains. Since conducting hundreds of radiocarbon dates may be impractical and expensive to solve stratigraphic problems, deamidation levels obtained during routine ZooMS screening of faunal samples may help place them in a relative timeframe. For example, use the average deamidation value of bones to provide the relative age for every spit in the backdirt. Combining with ZooMS-identified fauna abundance and lithic characteristics of each spit, it ultimately aims to achieve the reconstruction of the original stratigraphy.

5. Final remarks

5.1 Conclusions

My doctoral research, presented here in the form of a series of accepted and submitted manuscripts, had three main outcomes:

- Improvement on the analytical workflow of ZooMS

The first part of my doctoral work was the refinement of an efficient workflow for high-throughput screening on fauna assemblages using ZooMS. The evaluation of four protocols on 400 samples from seven sites suggests the Acid-insoluble protocol with HCl demineralisation can yield the most collagen, while the AmBic protocol is adequate for well-preserved samples (Manuscript A). Both protocols are compatible with high-throughput analysis for large numbers of samples. Then, the introduction of a semi-automated data processing pipeline, based on QuickID, enables rapid ZooMS spectra processing with minimal training or coding requirements. This pipeline also offers consistent identification standards and improves data reproducibility.

- The application of the refined workflow on material from Europe and East Asia

The most extensive part of my doctoral research was two ZooMS screening projects. The first focused on analysing bone fragments from the backdirt excavation of Vogelherd, a cave in the Lone Valley of the Swabian Jura. ZooMS analysis identified three previously unknown human bones from 287 bone fragments. Moreover, in this project, I reviewed the evolutionary trajectory of zooarchaeology through the lens of the Vogelherd case study since its first excavation in 1931 (Table 3 of Manuscript C, page 119).

The second project, conducted at Jinsitai Cave on the eastern margin of the Mongolia Plateau, represents the first and largest application of ZooMS on Palaeolithic material in East Asia. Through this ZooMS analysis of 745 bones, I discovered two new taxa, camel and ostrich that were not previously identified in the zooarchaeological assemblage. Radiocarbon dating on five camel bones unveils a previously undetected, long-term pattern of camel exploitation at Jinsitai (Manuscript B).

Additionally, feasibility studies on four other Palaeolithic sites in China, which involved nearly 600 samples, indicate that the application of ZooMS in Palaeolithic fauna assemblages is promising for both caves and open-air sites.

- The contextualisation of ZooMS results and their combination with traditional zooarchaeological data

Continuing from the aforementioned ZooMS applications, I highlighted the crucial need for integrating ZooMS and zooarchaeology based on a comprehensive understanding of both methodologies. I introduced an extensive list of essential considerations for integration within the Palaeolithic context, including assemblage properties, reference preparation, taxonomic identification, and data interpretation (Manuscript C). This preliminary framework can serve as a resource for ZooMS and zooarchaeological analysts, aiding them in discerning the distinctive features of their fauna assemblages and helping in formulating research questions for a more thorough exploration of integration.

In summary, the assessment of ZooMS chemical protocols by 400 samples laid the foundation for a high-throughput ZooMS workflow, complemented by a QuickID-based identification pipeline for efficient data processing. This ZooMS workflow was then applied in Europe and extended to East Asia, involving the analysis of approximately 1600 samples. The results included the discovery of three new human fossils at Vogelherd, the revelation of two previously unidentified taxa at Jinsitai, and a basic understanding of collagen preservation of Palaeolithic contexts in East Asia. The integration of ZooMS data and traditional zooarchaeological results led to the development of a preliminary framework for comprehensive analysis.

5.2 Future directions

5.2.1 Additional screening projects in East Asia

Advancing our understanding of human evolution in East Asia necessitates more hominin fossils. The pilot studies conducted in key regions for human evolution in China, such as the Nihewan Basin and the Three Gorges, underscore the potential of ZooMS. The Nihewan Basin, often referred to as the "Olduvai of China” for the studies

of Quaternary geology, palaeontology, and Palaeolithic archaeology, encompasses over 400 Palaeolithic archaeological localities. The ZooMS identification rate managed to reach 60% in the open-air site Xibaimaying (ca. 50-30 ka) in Nihewan Basin. The Three Gorges, another significant region for human evolution, preserves over 100 Palaeolithic sites, particularly limestone caves that offer a favourable environment for collagen preservation. The analysis at Yumidong Cave in Three Gorges yielded a ZooMS identification rate of surprisingly 83%.

The widespread adoption of the wet/dry sediment sieving technique in new Palaeolithic campaigns in China facilitates parallel ZooMS analysis on sieved bone fragments alongside traditional zooarchaeological examination on plotted bone remains. Additionally, emphasis should be placed on the scrutiny of unidentifiable portions within old fauna collections. High-throughput ZooMS analysis, aided by laboratory robots (Oldfield et al., 2023) and (semi-) automated data processing pipelines, holds promise for the extensive integration of ZooMS and zooarchaeology.

5.2.2 Methodological protocol improvements

Several approaches are available for measuring peptide mass, with MALDI-TOF-MS offering the advantage of fast detection speed, albeit at the cost of detecting resolution and tandem capability with chromatography instruments. MALDI-FTICR-MS, on the other hand, is not only swift but also capable of achieving high resolution (Bray et al., 2023). It can directly quantify PMFs such as glutamine deamidation. The shared MALDI ionisation allows FTICR spectra to utilise the reference library proved by TOF. However, MALDI-FTICR-MS instruments are less common than MALDI-TOF-MS and considerably more expensive. Some researchers suggested replacing ZooMS with LC-MS/MS-based proteomic sequencing for large-scale taxonomic identifications (Mylopotamitaki et al., 2023; Rüther et al., 2022). While the method offers clear advantages in detecting low-abundance proteins, its incomplete database and higher costs (Mylopotamitaki et al., 2024) prohibit such replacement shortly. However, involving other instruments for ZooMS method development is advisable. For instance, exploring the offline combination of HPLC and MALDI-TOF/TOF-MS for ZooMS marker development is possible. It can overcome the limitations in routine marker development workflow that stem from different ionisation methods between MALDI-TOF-MS and LC-MS/MS.

Moreover, improvements can be made to the eco-friendliness of the ZooMS extraction protocol. For well-preserved samples that yields detectable COL1 peptides with >20 times dilution, the purification step for buffer changing by C18 ZipTips may be unnecessary, as the dilution makes the buffer salt negligible. Additionally, HCl is no longer the exclusive demineralisation buffer. In a recent publication, EDTA was used in a “one-pot” extraction and proved compatible with trypsin digestion (Jensen et al., 2023), although the efficiency of EDTA demineralisation has not yet been compared with that of HCl. Ultrasound-assisted digestion presents a viable alternative by reducing the trypsin digestion time of cell lysate to approximately three hours (Huang et al., 2023). This method results in energy savings from the overnight incubation for COL1 digestion in ZooMS protocols, which is also worth exploring.

5.2.3 Expansion of the reference database

A local reference library is an important tool to improve the resolution of ZooMS results. While the limited ZooMS resolution in Europe and northern Eurasia projects is usually attributed to the conserved nature of COL1, projects in East Asia face additional constraints due to the absence of local taxa in the reference library.

Although well-known East Asian species, such as the giant panda and musk deer, are already included in the current ZooMS reference, some lesser-known and extinct species are not. For instance, in the ZooMS screening project at Jinsitai, the local marmot and extinct ostrich even lacked comparable ZooMS reference of their congener. Also at Jinsitai, local Antilopinae taxa constituted 12% of the assemblage, represented by four combinations of ZooMS markers (each consisting of seven markers), reflecting at least four ZooMS distinguishable taxa. The lack of reference development work on local taxa made associating specific marker combinations with particular species challenging.

Recent efforts in Australia and Africa for marker development and COL1 *de-novo* sequencing on local mammals have expanded the region for ZooMS applications (Janzen et al., 2021; Le Meillour et al., 2023; Peters et al., 2021). Future endeavours should focus on the development of reference markers for more East Asian species. Minimal-invasive ZooMS sampling techniques, such as using sandpaper (Chen, 2023)

and dermatological skin tape-disc (Fabrizi et al., 2023), also make it easier for museum curators and zooarchaeologists to allow the sampling of skeleton collections.

Reference

Bae, C.J., Douka, K., Petraglia, M.D., 2017. On the origin of modern humans: Asian perspectives. Science 358. https://doi.org/10.1126/science.aai9067

Bae, C.J., Wang, W., Zhao, J., Huang, S., Tian, F., Shen, G., 2014. Modern human teeth from Late Pleistocene Luna Cave (Guangxi, China). Quat. Int. 354, 169–183.

Baker, A., Harvey, V.L., Buckley, M., 2023. Machine Learning for collagen peptide biomarker determination in the taxonomic identification of archaeological fish remains. J. Archaeol. Sci. Rep. 49, 104001.

Bar-Yosef, O., Eren, M.I., Yuan, J., Cohen, D.J., Li, Y., 2012. Were bamboo tools made in prehistoric Southeast Asia? An experimental view from South China. Quat. Int. 269, 9–21.

Bednarik, R.G., 2011. About ostrich eggshell beads, in: The Bead Forum. pp. 2–8.

Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E.M.L., Skoglund, P., 2021. Origins of modern human ancestry. Nature 590, 229–237.

Boger, U., Starkovich, B.M., Conard, N.J., 2014. New insights gained from the faunal material recovered during the latest excavations at Vogelherd Cave. Mitteilungen der Gesellschaft für Urgeschichte 23, 57–81.

Brandt, L.Ø., Ebsen, J.A., Haase, K., 2020. Leather Shoes in Early Danish Cities: Choices of Animal Resources and Specialization of Crafts in Viking and Medieval Denmark. European Journal of Archaeology 23, 428–450.

Brandt, L.Ø., Mannering, U., 2021. Taxonomic identification of Danish Viking Age shoes and skin objects by ZooMS (Zooarchaeology by mass spectrometry). J. Proteomics 231, 104038.

Bray, F., Fabrizi, I., Flament, S., Locht, J.-L., Antoine, P., Auguste, P., Rolando, C., 2023. Robust high-throughput proteomics identification and deamidation quantitation of extinct species up to Pleistocene with ultrahigh-resolution MALDI-FTICR mass spectrometry. Anal. Chem. 95, 7422–7432.

Brock, F., Higham, T., Ditchfield, P., Ramsey, C.B., 2010. Current Pretreatment Methods for AMS Radiocarbon Dating at the Oxford Radiocarbon Accelerator Unit (Orau). Radiocarbon 52, 103–112.

Brock, F., Wood, R., Higham, T.F.G., Ditchfield, P., Bayliss, A., Ramsey, C.B., 2012.

Reliability of Nitrogen Content (%N) and Carbon: Nitrogen Atomic Ratios (C:N) as Indicators of Collagen Preservation Suitable for Radiocarbon Dating. Radiocarbon 54, 879–886.

Brown, S., 2021. Identifying ZooMS Spectra (mammals) using mMass. https://dx.doi.org/10.17504/protocols.io.bzscp6aw

Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., Derevianko, A., Buckley, M., 2016. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Sci. Rep. 6, 23559.

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

Brown, S., Kozlikin, M., Shunkov, M., Derevianko, A., Higham, T., Douka, K., Richter, K.K., 2021b. Examining collagen preservation through glutamine deamidation at Denisova Cave. J. Archaeol. Sci. 133, 105454.

Brown, S., Massilani, D., Kozlikin, M.B., Shunkov, M.V., Derevianko, A.P., Stoessel, A., Jope-Street, B., Meyer, M., Kelso, J., Pääbo, S., Higham, T., Douka, K., 2021c. The earliest Denisovans and their cultural adaptation. Nature Ecology & Evolution 6, 28–35.

Brown, S., Wang, N., Oertle, A., Kozlikin, M.B., Shunkov, M.V., Derevianko, A.P., Comeskey, D., Jope-Street, B., Harvey, V.L., Chowdhury, M.P., Buckley, M., Higham, T., Douka, K., 2021d. Zooarchaeology through the lens of collagen fingerprinting at Denisova Cave. Sci. Rep. 11, 15457.

Buckley, M., 2018. Zooarchaeology by Mass Spectrometry (ZooMS) Collagen Fingerprinting for the Species Identification of Archaeological Bone Fragments. In: Giovas, C.M., LeFebvre, M.J. (Eds.), Zooarchaeology in Practice: Case Studies in Methodology and Interpretation in Archaeofaunal Analysis. Springer International Publishing, Cham. pp. 227–247.

Buckley, M., Collins, M., Thomas-Oates, J., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun. Mass Sp. 23, 3842-3854.

Buckley, M., Harvey, V.L., Chamberlain, A.T., 2017. Species identification and decay assessment of Late Pleistocene fragmentary vertebrate remains from Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting. Boreas 46, 402–411.

Buckley, M., Whitcher Kansa, S., Howard, S., Campbell, S., Thomas-Oates, J.,

Collins, M., 2010. Distinguishing between archaeological sheep and goat bones using a single collagen peptide. J. Archaeol. Sci. 37, 13–20.

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

Chen, P., 2023. Comparison of Sandpaper and Polishing Film in Minimally-Invasive ZooMS. The Ethnograph 7, 48–57.

Cleland, T.P., 2018. Human Bone Paleoproteomics Utilizing the Single-Pot, Solid-Phase-Enhanced Sample Preparation Method to Maximize Detected Proteins and Reduce Humics. J. Proteome Res. 17, 3976–3983.

Cleland, T.P., Voegele, K., Schweitzer, M.H., 2012. Empirical evaluation of bone extraction protocols. PLoS One 7, e31443.

Codlin, M.C., Douka, K., Richter, K.K., 2022. An application of zooms to identify archaeological avian fauna from Teotihuacan, Mexico. J. Archaeol. Sci. 148, 105692.

Collins, M.J., Galley, P., 1998. Towards an optimal method of archaeological collagen extraction: the influence of pH and grinding. Anc. Biomol. 2, 209-223.

Conard, N.J., 2009. Jünger als gedacht! Zur Neudatierung der Menschenreste vom Vogelherd. In: Archäologisches Landesmuseum Baden-Württemberg/Abteilung Ältere Urgeschichte und Quartärökologie der Eberhard Karls Universität Tübingen (Eds.), Begleitband zur Großen Landesausstellung Eiszeit – Kunst und Kultur im Kunstgebäude Stuttgart. Jan Thorbecke Verlag, Ostfildern, pp. 116.

Conard, N.J., Grootes, P.M., Smith, F.H., 2004. Unexpectedly recent dates for human remains from Vogelherd. Nature. 430, 198–201.

Conard, N.J., Zeidi, M., Bega, J., 2012. Die letzte Kampagne der Nachgrabungen am Vogelherd. Archäol. Ausgrabungen Baden-Württemberg, 2012, 84-88.

Coutu, A.N., Whitelaw, G., le Roux, P., Sealy, J., 2016. Earliest Evidence for the Ivory Trade in Southern Africa: Isotopic and ZooMS Analysis of Seventh–Tenth Century ad Ivory from KwaZulu-Natal. Afr. Archaeol. Review 33, 411–435.

Creecy, A., Brown, K.L., Rose, K.L., Voziyan, P., Nyman, J.S., 2021. Post-translational modifications in collagen type I of bone in a mouse model of

aging. Bone. 143, 115763.

Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B.M., Rakownikow Jersie-Christensen, R., Olsen, J.V., Haile, J., Thomas, J., Marean, C.W., Parkington, J., Presslee, S., Lee-Thorp, J., Ditchfield, P., Hamilton, J.F., Ward, M.W., Wang, C.M., Shaw, M.D., Harrison, T., Domínguez-Rodrigo, M., MacPhee, R.D.E., Kwekason, A., Ecker, M., Kolska Horwitz, L., Chazan, M., Kröger, R., Thomas-Oates, J., Harding, J.H., Cappellini, E., Penkman, K., Collins, M.J., 2016. Protein sequences bound to mineral surfaces persist into deep time. Elife 5. https://doi.org/10.7554/eLife.17092

Derevianko, A.P., Rybin, E.P., 2003. The earliest representations of symbolic behavior by Paleolithic humans in the Altai Mountains. Archaeol. Ethnol. Anthro. Eurasia. 3, 27–50.

Devièse, T., Karavanić, I., Comeskey, D., Kubiak, C., Korlević, P., Hajdinjak, M., Radović, S., Procopio, N., Buckley, M., Pääbo, S., Higham, T., 2017. Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition. Proc. Natl. Acad. Sci. U. S. A. 114, 10606–10611.

Dobberstein, R.C., Collins, M.J., Craig, O.E., Taylor, G., Penkman, K.E.H., Ritz-Timme, S., 2009. Archaeological collagen: Why worry about collagen diagenesis? Archaeol. Anthropol. Sci. 1, 31–42.

Dong, W., Hou, Y.-M., Yang, Z.-M., Zhang, L.-M., Zhang, S.-Q., Liu, Y., 2014. Late Pleistocene mammalian fauna from Wulanmulan Paleolithic Site, Nei Mongol, China. Quat. Int. 347, 139–147.

Driver, J.C., Bovy, K., Butler, V.L., Lupo, K.D., Lyman, L.R., Otaola, C., 2011. Identification, Classification and Zooarchaeology. Ethnobiology Letters 2, 19–39.

Eda, M., Morimoto, M., Mizuta, T., Inoué, T., 2020. ZooMS for birds: Discrimination of Japanese archaeological chickens and indigenous pheasants using collagen peptide fingerprinting. Journal of Archaeological Science: Reports 34, 102635.

Fabrizi, I., Flament, S., Delhon, C., Gourichon, L., Vuillien, M., Oueslati, T., Auguste, P., Rolando, C., Bray, F., 2023. Low-invasive sampling method for taxonomic for the identification of archaeological and paleontological bones by proteomics of their collagens. bioRxiv. https://doi.org/10.1101/2023.10.18.562897

Fiddyment, S., Holsinger, B., Ruzzier, C., Devine, A., Binois, A., Albarella, U., Fischer, R., Nichols, E., Curtis, A., Cheese, E., Teasdale, M.D., Checkley-Scott, C., Milner, S.J., Rudy, K.M., Johnson, E.J., Vnouček, J., Garrison, M., McGrory, S., Bradley, D.G., Collins, M.J., 2015. Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. Proc. Natl. Acad. Sci. U. S. A. 112, 15066–15071.

Gladyshev, S.A., Olsen, J.W., Tabarev, A.V., Jull, A.J.T., 2012. The Upper Paleolithic of Mongolia: Recent finds and new perspectives. Quat. Int. 281, 36–46.

Gladyshev, S.A., Olsen, J.W., Tabarev, A.V., Kuzmin, Y.V., 2010. Chronology and periodization of upper paleolithic sites in Mongolia. Archaeol. Ethnol. Anthr. Eurasia 38, 33–40.

Groucutt, H.S., Grün, R., Zalmout, I.A.S., Drake, N.A., Armitage, S.J., Candy, I., Clark-Wilson, R., Louys, J., Breeze, P.S., Duval, M., Buck, L.T., Kivell, T.L., Pomeroy, E., Stephens, N.B., Stock, J.T., Stewart, M., Price, G.J., Kinsley, L., Sung, W.W., Alsharekh, A., Al-Omari, A., Zahir, M., Memesh, A.M., Abdulshakoor, A.J., Al-Masari, A.M., Bahameem, A.A., Al Murayyi, K.M.S., Zahrani, B., Scerri, E.L.M., Petraglia, M.D., 2018. Homo sapiens in Arabia by 85,000 years ago. Nat Ecol Evol 2, 800–809.

Gu, M., Buckley, M., 2018. Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. BMC Bioinformatics 19, 241.

Hahn, J., 1988. Die Geißenklösterle-Höhle im Achtal bei Blaubeuren I: Fundhorizontbildung und Besiedlung im Mittelpaläolithikum und im Aurignacien. Theiss, Stuttgart.

Hahn, J., 1977. Aurignacien, das ältere Jungpaläolithikum in Mittel-und Osteuropa. Böhlau, Cologne.

Han, F., Bahain, J.-J., Deng, C., Boëda, É., Hou, Y., Wei, G., Huang, W., Garcia, T., Shao, Q., He, C., Falguères, C., Voinchet, P., Yin, G., 2017. The earliest evidence of hominid settlement in China: Combined electron spin resonance and uranium series (ESR/U-series) dating of mammalian fossil teeth from Longgupo cave. Quat. Int. 434, 75–83.

Harvati, K., Röding, C., Bosman, A.M., Karakostis, F.A., Grün, R., Stringer, C., Karkanas, P., Thompson, N.C., Koutoulidis, V., Moulopoulos, L.A., Gorgoulis,

V.G., Kouloukoussa, M., 2019. Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia. Nature 571, 500–504.

Harvey, V.L., LeFebvre, M.J., Sharpe, A.E., Toftgaard, C., deFrance, S.D., Giovas, C.M., Fitzpatrick, S.M., Buckley, M., 2022. Collagen fingerprinting of Caribbean archaeological fish bones: Methodological implications for historical fisheries baselines and anthropogenic change. J. Archaeol. Sci. 145, 105642.

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

Huang, Y., Shao, X., Liu, Y., Yan, K., Ying, W., He, F., Wang, D., 2023. RUPE-phospho: Rapid Ultrasound-Assisted Peptide-Identification-Enhanced Phosphoproteomics Workflow for Microscale Samples. Anal. Chem. 49, 17974-17980. https://doi.org/10.1021/acs.analchem.3c02623

Hublin, J.-J., Sirakov, N., Aldeias, V., Bailey, S., Bard, E., Delvigne, V., Endarova, E., Fagault, Y., Fewlass, H., Hajdinjak, M., Kromer, B., Krumov, I., Marreiros, J., Martisius, N.L., Paskulin, L., Sinet-Mathiot, V., Meyer, M., Pääbo, S., Popov, V., Rezek, Z., Sirakova, S., Skinner, M.M., Smith, G.M., Spasov, R., Talamo, S., Tuna, T., Wacker, L., Welker, F., Wilcke, A., Zahariev, N., McPherron, S.P., Tsanova, T., 2020. Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. Nature 581, 299–302.

Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., Krijgsveld, J., 2014. Ultrasensitive proteome analysis using paramagnetic bead technology. Mol. Syst. Biol. 10, 757.

Jain, S., Rai, N., Kumar, G., Pruthi, P.A., Thangaraj, K., Bajpai, S., Pruthi, V., 2017. Ancient DNA Reveals Late Pleistocene Existence of Ostriches in Indian Sub-Continent. PLoS One 12, e0164823.

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

Jensen, T.Z.T., Yeomans, L., Le Meillour, L., Nielsen, P.W., Ramsøe, M., Mackie, M., Bangsgaard, P., Kinzel, M., Thuesen, I., Collins, M., Taurozzi, A.J., 2023. Tryps-In: A Streamlined Palaeoproteomics Workflow Enables Zooms Analysis

of 10,000-Year-Old Petrous Bones from Jordan Rift-Valley. https://doi.org/10.2139/ssrn.4441624

Khatsenovich, A.M., Rybin, E.P., Tserendagva, Y., Bazargur, D., Margad-Erdene, G., Marchenko, D.V., Gunchinsuren, B., Olsen, J.W., Derevianko, A.P., 2023. The Middle Paleolithic of Tsagaan Agui Cave in the Gobi Altai region of Mongolia and its Siberian and Central Asian links. Archaeological Research in Asia 35, 100462.

Klementiev, A.M., Khatsenovich, A.M., Tserendagva, Y., Rybin, E.P., Bazargur, D., Marchenko, D.V., Gunchinsuren, B., Derevianko, A.P., Olsen, J.W., 2022. First Documented Camelus knoblochi Nehring (1901) and Fossil Camelus ferus Przewalski (1878) From Late Pleistocene Archaeological Contexts in Mongolia. Front. Earth Sci. 10. https://doi.org/10.3389/feart.2022.861163

Kurochkin, E.N., Kuzmin, Y.V., Antoshchenko-Olenev, I.V., Zabelin, V.I., Krivonogov, S.K., Nohrina, T.I., Lbova, L.V., Burr, G.S., Cruz, R.J., 2010. The timing of ostrich existence in Central Asia: AMS 14C age of eggshells from Mongolia and southern Siberia (a pilot study). Nucl. Instrum. Methods Phys. Res. B 268, 1091–1093.

Lehmann, U., 1954. Die Fauna des “Vogelherds” bei Stetten ob Lontal (Württemberg). Neues Jahrb. Geol. Paläontol 99: 33-146.

Le Meillour, L., Zazzo, A., Zirah, S., Tombret, O., Barriel, V., Arthur, K.W., Arthur, J.W., Cauliez, J., Chaix, L., Curtis, M.C., Gifford-Gonzalez, D., Gunn, I., Gutherz, X., Hildebrand, E., Khalidi, L., Millet, M., Mitchell, P., Studer, J., Vila, E., Welker, F., Pleurdeau, D., Lesur, J., 2023. The name of the game: palaeoproteomics and radiocarbon dates further refine the presence and dispersal of caprines in eastern and southern Africa. R Soc Open Sci 10, 231002.

Liao, W., Xing, S., Li, D., Martinón-Torres, M., Wu, X., Soligo, C., Bermúdez de Castro, J.M., Wang, W., Liu, W., 2019. Mosaic dental morphology in a terminal Pleistocene hominin from Dushan Cave in southern China. Sci. Rep. 9, 2347.

Li, F., Kuhn, S.L., Chen, F., Wang, Y., Southon, J., Peng, F., Shan, M., Wang, C., Ge, J., Wang, X., Yun, T., Gao, X., 2018. The easternmost Middle Paleolithic (Mousterian) from Jinsitai Cave, North China. J. Hum. Evol. 114, 76–84.

Li, F., Vanwezer, N., Boivin, N., Gao, X., Ott, F., Petraglia, M., Roberts, P., 2019. Heading north: Late Pleistocene environments and human dispersals in central

and eastern Asia. PLoS One 14, e0216433.

Liu, W., Martinón-Torres, M., Cai, Y.-J., Xing, S., Tong, H.-W., Pei, S.-W., Sier, M.J., Wu, X.-H., Edwards, R.L., Cheng, H., Li, Y.-Y., Yang, X.-X., de Castro, J.M.B., Wu, X.-J., 2015. The earliest unequivocally modern humans in southern China. Nature 526, 696–699.

Longin, R., 1971. New method of collagen extraction for radiocarbon dating. Nature 230, 241–242.

Lyman, L.R., 2018. Observations on the history of zooarchaeological quantitative units: Why NISP, then MNI, then NISP again? Journal of Archaeological Science: Reports 18, 43–50.

Martisius, N.L., Welker, F., Dogandžić, T., Grote, M.N., Rendu, W., Sinet-Mathiot, V., Wilcke, A., McPherron, S.J.P., Soressi, M., Steele, T.E., 2020. Non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals. Sci. Rep. 10, 7746.

Martynovich, N., 2002. Pleistocene birds from Tsagan-Agui Cave (Gobian Altai). Acta Zool. Cracov. 45, 283.

Miller, J.M., Wang, Y.V., 2022. Ostrich eggshell beads reveal 50,000-year-old social network in Africa. Nature 601, 234–239.

Morales, J.I., Cebrià, A., Vergès, J.M., Bañuls-Cardona, S., Cervelló, J.M., Hernando, R., Lombao, D., Marín, J., Marsal, R., Xavier Oms, F., Rabuñal, J., Rodríguez-Hidalgo, A., Soto, M., Rosas, A., Fullola, J.M., 2022. Palaeolithic archaeology in the conglomerate caves of north-eastern Iberia. Antiquity 96, 1–9.

Morin, E., Oldfield, E.-M., Baković, M., Bordes, J.-G., Castel, J.-C., Crevecoeur, I., Rougier, H., Monnier, G., Tostevin, G., Buckley, M., 2023. A double-blind comparison of morphological and collagen fingerprinting (ZooMS) methods of skeletal identifications from Paleolithic contexts. Sci. Rep. 13, 18825.

Mylopotamitaki, D., Harking, F.S., Taurozzi, A.J., Fagernäs, Z., Godinho, R.M., Smith, G.M., Weiss, M., Schüler, T., McPherron, S.P., Meller, H., Cascalheira, J., Bicho, N., Olsen, J.V., Hublin, J.-J., Welker, F., 2023. Comparing extraction method efficiency for high-throughput palaeoproteomic bone species identification. Sci. Rep. 13, 18345.

Mylopotamitaki, D., Weiss, M., Fewlass, H., Zavala, E.I., Rougier, H., Sümer, A.P., Hajdinjak, M., Smith, G.M., Ruebens, K., Sinet-Mathiot, V., Pederzani, S.,

Essel, E., Harking, F.S., Xia, H., Hansen, J., Kirchner, A., Lauer, T., Stahlschmidt, M., Hein, M., Talamo, S., Wacker, L., Meller, H., Dietl, H., Orschiedt, J., Olsen, J.V., Zeberg, H., Prüfer, K., Krause, J., Meyer, M., Welker, F., McPherron, S.P., Schüler, T., Hublin, J.-J., 2024. Homo sapiens reached the higher latitudes of Europe by 45,000 years ago. Nature. https://doi.org/10.1038/s41586-023-06923-7

Nejman, L., Hughes, P., Sullivan, M., Wright, D., Way, A.M., Skopal, N., Mlejnek, O., Škrdla, P., Lisá, L., Kmošek, M., Nývltová Fišáková, M., Králík, M., Neruda, P., Nerudová, Z., Přichystal, A., 2020. Preliminary report of the 2019 excavation at Švédův Stůl Cave in the Moravian Karst. Přehled výzkumů 11–19.

Niedermeyer, T.H.J., Strohalm, M., 2012. mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. PLoS One 7, e44913.

Niven, L., 2007. From carcass to cave: large mammal exploitation during the Aurignacian at Vogelherd, Germany. J. Hum. Evol. 53, 362–382.

Niven, L., 2006. The palaeolithic occupation of Vogelherd cave: implications for the subsistence behavior of late Neanderthals and early modern humans. Kerns, Tübingen.

Niven, L., 2001. The role of mammoths in Upper Palaeolithic economies of southern Germany. In: Cavarretta, G., Gioia, P., Mussi, M., Palombo, M. (Eds.), The World of Elephants: Proceedings of the 1st International Congress. CNR, Rome, pp. 323-327.

Ni, X., Ji, Q., Wu, W., Shao, Q., Ji, Y., Zhang, C., Liang, L., Ge, J., Guo, Z., Li, J., Li, Q., Grün, R., Stringer, C., 2021. Massive cranium from Harbin in northeastern China establishes a new Middle Pleistocene human lineage. Innovation (Camb) 2, 100130.

Oldfield, E.-M., Dunstan, M., Chowdhury, M.P., Slimak, L., Buckley, M., 2023. AutoZooMS: Integrating robotics into high-throughput ZooMS for the species identification of archaeofaunal remains at Grotte Mandrin, France. Research Square. https://doi.org/10.21203/rs.3.rs-2762261/v1

Paladugu, R., Richter, K.K., Valente, M.J., Gabriel, S., Detry, C., Warinner, C., Dias, C.B., 2023. Your horse is a donkey! Identifying domesticated equids from Western Iberia using collagen fingerprinting. J. Archaeol. Sci. 149, 105696.

Pei, S., Gao, X., Wu, X., Li, X., Bae, C.J., 2013. Middle to Late Pleistocene hominin occupation in the Three Gorges region, South China. Quat. Int. 295, 237–252.

Peters, C., Richter, K.K., Manne, T., Dortch, J., Paterson, A., Travouillon, K., Louys, J., Price, G.J., Petraglia, M., Crowther, A., Boivin, N., 2021. Species identification of Australian marsupials using collagen fingerprinting. R Soc Open Sci 8, 211229.

Pirson, S., Di Modica, K., Jungels, C., Flas, D., Hauzeur, A., Toussaint, M., 2012. The stratigraphy of spy cave: A review of the available lithostratigraphic and archaeostratigraphic information. Anthropol. Praehist. 123, 91–131.

Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., Francken, M., Malina, M., Bolus, M., Lari, M., Gigli, E., Capecchi, G., Crevecoeur, I., Beauval, C., Flas, D., Germonpré, M., van der Plicht, J., Cottiaux, R., Gély, B., Ronchitelli, A., Wehrberger, K., Grigorescu, D., Svoboda, J., Semal, P., Caramelli, D., Bocherens, H., Harvati, K., Conard, N.J., Haak, W., Powell, A., Krause, J., 2016. Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. Curr. Biol. 26, 827–833.

Pothier Bouchard, G., Riel-Salvatore, J., Negrino, F., Buckley, M., 2020. Archaeozoological, taphonomic and ZooMS insights into The Protoaurignacian faunal record from Riparo Bombrini. Quat. Int. 551, 243–263.

Qi, G., 1975. Quaternary mammalian fossils from Salawusu river district, Nei Mongol. Vertebrata Palasiatica. 13, 239-249.

Richter, K.K., Wilson, J., Jones, A.K.G., Buckley, M., van Doorn, N., Collins, M.J., 2011. Fish ’n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments. J. Archaeol. Sci. 38, 1502–1510.

Riek, G., 1934. Die Eiszeitjägerstation am Vogelherd im Lonetal. Akademische Verlagsbuchhandlung Franz F. Heine, Tübingen.

Ruebens, K., Sinet-Mathiot, V., Talamo, S., Smith, G.M., Welker, F., Hublin, J.-J., McPherron, S.P., 2022. The Late Middle Palaeolithic Occupation of Abri du Maras (Layer 1, Neronian, Southeast France): Integrating Lithic Analyses, ZooMS and Radiocarbon Dating to Reconstruct Neanderthal Hunting Behaviour. J. Paleolithic Archaeol. 5, 4.

Ruebens, K., Smith, G.M., Fewlass, H., Sinet-Mathiot, V., Hublin, J.-J., Welker, F., 2023. Neanderthal subsistence, taphonomy and chronology at Salzgitter-Lebenstedt (Germany): a multifaceted analysis of morphologically

unidentifiable bone. J. Quat. Sci. 38(4), 471-87.

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

Rybczynski, N., Gosse, J.C., Harington, C.R., Wogelius, R.A., Hidy, A.J., Buckley, M., 2013. Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. Nat. Commun. 4, 1550.

Rybin, E.P., Khatsenovich, A.M., 2020. Middle and Upper Paleolithic Levallois technology in eastern Central Asia. Quat. Int. 535, 117–138.

Schmitz, R.W., Serre, D., Bonani, G., Feine, S., Hillgruber, F., Krainitzki, H., Pääbo, S., Smith, F.H., 2002. The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. Proc. Natl. Acad. Sci. U. S. A. 99, 13342–13347.

Shackelford, L., Demeter, F., Westaway, K., Duringer, P., Ponche, J.-L., Sayavongkhamdy, T., Zhao, J.-X., Barnes, L., Boyon, M., Sichanthongtip, P., Sénégas, F., Patole-Edoumba, E., Coppens, Y., Dumoncel, J., Bacon, A.-M., 2018. Additional evidence for early modern human morphological diversity in Southeast Asia at Tam Pa Ling, Laos. Quat. Int. 466, 93–106.

Shao, Q., Philippe, A., He, C., Jin, M., Huang, M., Jiao, Y., Voinchet, P., Lin, M., Bahain, J.-J., 2022. Applying a Bayesian approach for refining the chronostratigraphy of the Yumidong site in the Three Gorges region, central China. Quat. Geochronol. 70, 101304.

Shen, G., Wu, X., Wang, Q., Tu, H., Feng, Y.-X., Zhao, J.-X., 2013. Mass spectrometric U-series dating of Huanglong Cave in Hubei Province, Central China: evidence for early presence of modern humans in Eastern Asia. J. Hum. Evol. 65, 162–167.

Shoulders, M.D., Raines, R.T., 2009. Collagen structure and stability. Annu. Rev. Biochem. 78, 929–958.

Silvestrini, S., Lugli, F., Romandini, M., Real, C., Sommella, E., Salviati, E., Arrighi, S., Bortolini, E., Figus, C., Higgins, O.A., Marciani, G., Oxilia, G., Delpiano, D., Vazzana, A., Piperno, M., Crescenzi, C., Campiglia, P., Collina, C., Peresani, M., Spinapolice, E.E., Benazzi, S., 2022. Integrating ZooMS and

zooarchaeology: New data from the Uluzzian levels of Uluzzo C Rock Shelter, Roccia San Sebastiano cave and Riparo del Broion. PLoS One 17, e0275614.

Simpson, J.P., Penkman, K.E.H., Demarchi, B., Koon, H., Collins, M.J., Thomas-Oates, J., Shapiro, B., Stark, M., Wilson, J., 2016. The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. J. Archaeol. Sci. 69, 29–38.

Sinet-Mathiot, V., Rendu, W., Steele, T.E., Spasov, R., Madelaine, S., Renou, S., Soulier, M.-C., Martisius, N.L., Aldeias, V., Endarova, E., Goldberg, P., McPherron, S.J.P., Rezek, Z., Sandgathe, D., Sirakov, N., Sirakova, S., Soressi, M., Tsanova, T., Turq, A., Hublin, J.-J., Welker, F., Smith, G.M., 2023. Identifying the unidentified fauna enhances insights into hominin subsistence strategies during the Middle to Upper Palaeolithic transition. Archaeol. Anthropol. Sci. 15, 139.

Sinet-Mathiot, V., Smith, G.M., 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). Sci. Rep. 9, 12350.

Sjögren, K.-G., Buckley, M., Vretemark, M., Axelsson, T., 2023. Evaluating caprine remains of the Swedish Funnel Beaker culture through ZooMS. Archaeol. Anthropol. Sci. 15, 47.

Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M.B., Shunkov, M.V., Derevianko, A.P., Kelso, J., Meyer, M., Prüfer, K., Pääbo, S., 2018. The genome of the offspring of a Neanderthal mother and a Denisovan father. Nature 561, 113–116.

Song, Y., Cohen, D.J., Shi, J., 2022. Diachronic Change in the Utilization of Ostrich Eggshell at the Late Paleolithic Shizitan Site, North China. Front Earth Sci. Chin. 9. https://doi.org/10.3389/feart.2021.818554

Tang, Z., Liu, S., Lin, Z., Liu, H., 2003. The late Pleistocene fauna from Dabusu of Qian'an in Jilin Province. Vertebrata Palasiatica 41, 137.

Titov, V.V., 2008. Habitat conditions for Camelus knoblochi and factors in its extinction. Quat. Int. 179, 120-125.

Torres-Iglesias, L., Marín-Arroyo, A.B., Welker, F., 2024. Using ZooMS to assess archaeozoological insights and unravel human subsistence behaviour at La

Viña rock shelter (northern Iberia). J. Archaeol. Sci., 161, 105904.
van der Sluis, L.G., Hollund, H.I., Buckley, M., De Louw, P.G.B., Rijsdijk, K.F., Kars, H., 2014. Combining histology, stable isotope analysis and ZooMS collagen fingerprinting to investigate the taphonomic history and dietary behaviour of extinct giant tortoises from the Mare aux Songes deposit on Mauritius. Palaeogeogr. Palaeoclimatol. Palaeoecol. 416, 80–91.
van Doorn, N.L., Hollund, H., Collins, M.J., 2011. A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. Archaeol. Anthropol. Sci. 3, 281–289.
van Doorn, N.L., Wilson, J., Hollund, H., Soressi, M., Collins, M.J., 2012. Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. Rapid Commun. Mass Spectrom. 26, 2319–2327.
Velliky, E.C., Schmidt, P., Bellot-Gurlet, L., Wolf, S., Conard, N.J., 2021. Early anthropogenic use of hematite on Aurignacian ivory personal ornaments from Hohle Fels and Vogelherd caves, Germany. J. Hum. Evol. 150, 102900.
Wadsworth, C., Buckley, M., 2014. Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone. Rapid Commun. Mass Spectrom. 28, 605–615.
Wang, X., Wei, J., Chen, Q., Tang, Z., Wang, C., 2010. A preliminary study on the excavation of the Jinsitai cave site. Acta Anthropologica Sinica 29, 15–30.
Wang, N., Brown, S., Ditchfield, P., Hebestreit, S., Kozilikin, M., Luu, S., Wedage, O., Grimaldi, S., Chazan, M., Horwitz, H.L., Spriggs, M., Summerhayes, G., Shunkov, M., Richter, R.K., Douka, K., 2021. Testing the efficacy and comparability of ZooMS protocols on archaeological bone. J. Proteomics 233, 104078.
Wang, N., Xu, Y., Tang, Z., He, C., Hu, X., Cui, Y., Douka, K., 2023. Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China. Proc. Biol. Sci. 290, 20231129.
Weiner, S., Wagner, H.D., 1998. THE MATERIAL BONE: Structure-Mechanical Function Relations. Annu. Rev. Mater. Sci. 28, 271–298.
Wei, Y., d'Errico, F., Vanhaeren, M., Peng, F., Chen, F., Gao, X., 2017. A technological and morphological study of Late Paleolithic ostrich eggshell beads from Shuidonggou, North China. J. Archaeol. Sci. 85, 83–104.

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, B.M., Stewart, J.R., Pääbo, S., Collins, M.J., Hublin, J.-J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proc. Natl. Acad. Sci. U. S. A. 113, 11162–11167.

Welker, F., Smith, G.M., Hutson, J.M., Kindler, L., Garcia-Moreno, A., Villaluenga, A., Turner, E., Gaudzinski-Windheuser, S., 2017a. Middle Pleistocene protein sequences from the rhinoceros genus Stephanorhinus and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. PeerJ 5, 3033.

Welker, F., Soressi, M.A., Roussel, M., van Riemsdijk, I., Hublin, J.-J., Collins, M.J., 2017b. Variations in glutamine deamidation for a Châtelperronian bone assemblage as measured by peptide mass fingerprinting of collagen. STAR: Science & Technology of Archaeological Research 3, 15–27.

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. J. Archaeol. Sci. 54, 279–286.

Westaway, K.E., Louys, J., Due Awe, R., Morwood, M.J., Price, G.J., Zhao, J.-X., Aubert, M., Joannes-Boyau, R., Smith, T.M., Skinner, M.M., Compton, T., Bailey, R.M., van den Bergh, G.D., de Vos, J., Pike, A.W.G., Stringer, C., Saptomo, E.W., Rizal, Y., Zaim, J., Santoso, W.D., Trihascaryo, A., Kinsley, L., Sulistyanto, B., 2017. An early modern human presence in Sumatra 73,000–63,000 years ago. Nature 548, 322–325.

Wilson, J., van Doorn, N.L., Collins, M.J., 2012. Assessing the extent of bone degradation using glutamine deamidation in collagen. Anal. Chem. 84, 9041–9048.

Wolf, S., 2015. Personal ornaments as signatures of identity in the Aurignacian--the case of the Swabian Jura and western Germany, in: Human Origin Sites and the World Heritage Convention in Eurasia. UNESCO Publishing, pp. 92–102.

Wright, D., Hughes, P., Skopal, N., Kmošek, M., Way, A., Sullivan, M., Lisá, L., Ricardi, P., Škrdla, P., Nejman, L., Gadd, P., Fišáková, M.N., Mlejnek, O., Králík, M., 2021. The archaeology of overburden: Method within the madness at Švédův Stůl, Czech Republic. J. Archaeol. Sci. 132, 105429.

Yang, S.-X., Zhang, J.-F., Yue, J.-P., Wood, R., Guo, Y.-J., Wang, H., Luo, W.-G.,

Zhang, Y., Raguin, E., Zhao, K.-L., Zhang, Y.-X., Huan, F.-X., Hou, Y.-M., Huang, W.-W., Wang, Y.-R., Shi, J.-M., Yuan, B.-Y., Ollé, A., Queffelec, A., Zhou, L.-P., Deng, C.-L., d'Errico, F., Petraglia, M., 2024. Initial Upper Palaeolithic material culture by 45,000 years ago at Shiyu in northern China. Nat Ecol Evol. https://doi.org/10.1038/s41559-023-02294-4

Zhang, D., Xia, H., Chen, F., Li, B., Slon, V., Cheng, T., Yang, R., Jacobs, Z., Dai, Q., Massilani, D., Shen, X., Wang, J., Feng, X., Cao, P., Yang, M.A., Yao, J., Yang, J., Madsen, D.B., Han, Y., Ping, W., Liu, F., Perreault, C., Chen, X., Meyer, M., Kelso, J., Pääbo, S., Fu, Q., 2020. Denisovan DNA in Late Pleistocene sediments from Baishiya Karst Cave on the Tibetan Plateau. Science 370, 584–587.

Appendix 1 - Accepted publications

Manuscript A

Naihui Wang, Samantha Brown, Peter Ditchfield, Sandra Hebestreit, Maxim Kozilikin, Sindy Luu, Oshan Wedage, Stefano Grimaldi, Michael Chazan, Liora Kolska Horwitz, Matthew Spriggs, Glenn Summerhayes, Michael Shunkov, Kristine Korzow Richter, Katerina Douka. "Testing the efficacy and comparability of ZooMS protocols on archaeological bone." *Journal of Proteomics* 233 (2021): 104078.

Journal of Proteomics 233 (2021) 104078

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Testing the efficacy and comparability of ZooMS protocols on archaeological bone

Wang Naihui [a,*], Brown Samantha [a], Ditchfield Peter [b], Hebestreit Sandra [a], Kozilikin Maxim [c], Luu Sindy [d], Wedage Oshan [a,e], Grimaldi Stefano [f,g], Chazan Michael [h,i], Horwitz Kolska Liora [j], Spriggs Matthew [k], Summerhayes Glenn [l], Shunkov Michael [c], Richter Korzow Kristine [a,*], Douka Katerina [a,*]

[a] Department of Archaeology, Max Planck Institute for the Science of Human History (MPI-SHH), Kahlaische Straße 10, 07745 Jena, Germany
[b] School of Archaeology, University of Oxford, 1 South Parks Road, Oxford, UK
[c] Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Pr. Akademika Lavrentieva, 17, Novosibirsk 630090, Russia
[d] Department of Anatomy, University of Otago, PO Box 56, Dunedin 9054, New Zealand
[e] Department of History and Archaeology, University of Sri Jayewardenepura, Gangodawila, Nugegoda, Sri Lanka
[f] LaBAAF -Laboratorio Bagolini Archeologia, Archeometria, Fotografia, CeASUm – Centro di Alti Studi Umanistici, Dipartimento di Lettere e Filosofia, Università di Trento, via T.Gar14, I-38122 Trento, Italy
[g] IsIPU - Istituto Italiano di Paleontologia Umana, Anagni, Italy
[h] Department of Anthropology, Canada Institute of Evolutionary Studies, University of Toronto, 19 Russell Street, Toronto, Canada
[i] University of the Witwatersrand, Johannesburg, South Africa
[j] National Natural History Collections, The Hebrew University, Berman Building, E. Safra-Givat Ram Campus, 91904 Jerusalem, Israel
[k] School of Archaeology and Anthropology, Sir Roland Wilson Bldg 120, The Australian National University, Canberra, ACT 2600, Australia
[l] Archaeology Programme, Otago University, Dunedin, New Zealand.

ARTICLE INFO

Keywords:
Palaeoproteomics
ZooMS
Collagen
Bone
Nitrogen

ABSTRACT

Collagen peptide mass fingerprinting, best known as Zooarchaeology by Mass Spectrometry (or ZooMS) when applied to archaeology, has become invaluable for the taxonomic identification of archaeological collagenous materials, in particular fragmentary and modified bone remains. Prior to MALDI-based spectrometric analysis, collagen needs to be extracted from the bone's inorganic matrix, isolated and purified. Several protocols are currently employed for ZooMS analysis, however their efficacy and comparability has not been directly tested. Here, we use four different ZooMS protocols to analyze 400 bone samples from seven archaeological sites, dating to between ~500,000–2000 years ago. One of them, single-pot solid-phase-enhance sample preparation (SP3), is used for the first time as a ZooMS protocol. Our results indicate that the least-destructive ZooMS protocol which uses an ammonium bicarbonate buffer as a means of extracting collagen is most suitable for bones with good collagen preservation, whereas the acid-based methodologies can improve success rates for bones with low-to-medium collagen preservation. Since preservation of biomolecules in archaeological bones is highly variable due to age and environmental conditions, we use the percent nitrogen by weight (%N) value as an independent semi-quantitative proxy for assessing collagen content and for predicting which bones will likely result in a successful ZooMS-based identification. We find that 0.26%N as a threshold for screening material could optimize the number of spectra which produce identifications using ZooMS.

Significance statement: We present a direct comparison of three previously published ZooMS protocols for the analyses of archaeological bones, and the first use of an SP3-based approach to ZooMS analysis. Our results show that the acid-based ZooMS protocols increase the success rate for bones with low-medium collagen preservation. We identify 0.26%N as a threshold for optimizing the number of samples with enough collagen for successful peptide mass fingerprinting.

* Corresponding authors.
E-mail addresses: nwang@shh.mpg.de (W. Naihui), brown@shh.mpg.de (B. Samantha), peter.ditchfield@arch.ox.ac.uk (D. Peter), hebestreit@shh.mpg.de (H. Sandra), sindy.luu@postgrad.otago.ac.nz (L. Sindy), wedage@shh.mpg.de (W. Oshan), stefano.grimaldi@unitn.it (G. Stefano), mchazan@chass.utoronto.ca (C. Michael), Matthew.Spriggs@anu.edu.au (S. Matthew), glenn.summerhayes@otago.ac.nz (S. Glenn), kkrichter@shh.mpg.de (R.K. Kristine), douka@shh.mpg.de (D. Katerina).

https://doi.org/10.1016/j.jprot.2020.104078
Received 8 May 2020; Received in revised form 3 December 2020; Accepted 12 December 2020
Available online 15 December 2020

1. Introduction

Zooarchaeology by mass spectrometry (ZooMS), a form of peptide mass fingerprinting, has been used in palaeoproteomics for almost a decade as a quick and inexpensive way to taxonomically identify archaeological materials which cannot be determined morphologically [1]. ZooMS involves the extraction of collagen from a sample and the generation of tryptic-digested peptide mass fingerprints via MALDI-TOF-MS (Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry). Comparison of the fingerprints against a reference database of known – mostly mammalian – taxa allows identification to genus, and rarely to species level. Since its initial publication [1,2], ZooMS has been applied on a wide range of collagenous materials (e.g. leather, ivory, parchment) [3–6], although its broadest application has been on archaeological bone. It has been used to answer specific taxonomic questions [7], assess faunal assemblages [8,9], and determine human remains when these are too fragmentary to be identified otherwise [10–13].

Bones are one of the most important components of the archaeological and palaeontological records and are often abundant in excavated sites. They are a composite material of an organic fraction held within an inorganic matrix. Depending on the species and type of bone, living bone of land mammals is composed of 5–15% water, 20–35% organic components, and 60–70% inorganic components [14–16]. The vast majority of the organic component (~90%) is composed of the structural protein type I collagen (COL1). COL1, a fibrillar collagen, is a triple helix with each polypeptide chain around ~1000 amino acids long. In mammals, the helix is composed of two identical COL1α1 chains and one COL1α2 [17,18]. The inorganic component, largely calcium hydroxyapatite, forms a matrix that is located on the surface and within the collagen fibrils. Due to the large quantity of collagen in living bone and the intercalation of the mineral matrix with the collagen fibrils, bone collagen has been shown to have a longer preservation over other biomolecules, including DNA [19]. ZooMS has been applied to bone remains as old as 3.5 Ma [20] and collagen preservation limits are potentially much older.

COL1 is highly conserved with a slow evolutionary rate, probably due to thermal and functional constraints [21]. While this was originally considered a drawback for its use in phylogenetic separation, its analytical value was reconsidered on the basis of results stemming from peptide mass fingerprinting methodologies [1].

Peptide mass fingerprinting is useful for the identification of relatively homogenous samples, but becomes more complicated for the analyses of mixed proteins. Hence, sample preparation, which in the case of bones involves physical extraction of collagen from its inorganic matrix, purification and enzymatic digestion prior to MALDI-based mass spectrometric analysis, is a critical part of the workflow. Over the past years, several sample preparation protocols and variations of them have been introduced for efficient collagen extraction. In particular, a non-destructive protocol using ammonium bicarbonate (AmBic) without demineralization of the inorganic matrix was introduced in 2011 [22] and has gained popularity over the original acid-based protocol which involves demineralization of the inorganic bone matrix and subsequent analyses of the acid-soluble and acid-insoluble phases [7,23].

In this paper, we explore the efficacy of four different sample preparation protocols. In addition we assess a common screening tool for determining collagen preservation (%N content) and how this may relate to ZooMS efficiency.

1.1. ZooMS protocols

There are several variations of the protocols used to extract collagen from bones for ZooMS, generally falling into three categories:

(1) non-destructive analysis of the soluble collagen fraction without using acid [22];
(2) demineralization in acid followed by analysis of the acid insoluble fraction [1,2,24]; and.
(3) demineralization in acid followed by analysis of the acid soluble fraction [1,25].

The most widely used approach uses acid, usually hydrochloric acid (HCl), to demineralize 10–30 mg of the bone's inorganic bone matrix. After demineralization the acid is removed, the remaining acid-insoluble portion is washed with an ammonium bicarbonate buffer to increase the pH, and is heated to gelatinize the collagen. The supernatant is then digested with trypsin and the peptides are purified with C18 ZipTips [1,24]. We refer this method as the "acid-insoluble" protocol.

Following demineralization, the acid supernatant contains acid soluble collagen which can be extracted by filtering through a molecular-weight cutoff (MWCO) ultrafilter. This allows low-weight contaminants and inorganic salts to be removed and a buffer exchange to ammonium bicarbonate. The high molecular weight fraction, now in ammonium bicarbonate, is digested with trypsin and the peptides are then purified with C18 ZipTips. We refer to this method as the "acid-soluble" protocol.

In well preserved bones, good quality collagen is predominantly found in the HCl-insoluble fraction, whereas the HCl-soluble organic phase appears to be composed of the breakdown products of the insoluble collagen fraction [26]. As a result of the release of collagen trapped within the bone inorganic matrix, the acid protocols have been considered optimal for the ZooMS analyses of low-collagen bones. In some cases, e.g. in limited sample amounts or limited collagen yield, acid-insoluble and acid-soluble fractions may be combined before or after tryptic digestion [25].

Acid pretreatment is destructive to the bone and also may cause damage to the protein (specifically deamidation of glutamine and asparagine), hence it is not suitable for all samples. Specifically, valuable or small samples such as parchment, worked bone, or reference material often cannot be sub-sampled or destroyed for curatorial reasons. Therefore a non-destructive method was developed as an alternative to the acid-based approaches [22,23]. The demineralization step is excluded, and the sample is directly incubated in ammonium bicarbonate for 1–5 days followed by heating for an hour. It is then slowly dried out, while the supernatant undergoes digestion with trypsin and purification using C18 ZipTips. Since ammonium bicarbonate has no decalcifying potency, it is neither an acid nor a chelating agent, this step does not impact on collagen quality and is directly compatible with enzyme digestions [27]. For well-preserved samples, collagen from the bone surface and subsurface is assumed to leach and dissolve into the ammonium bicarbonate solution. The collagen fragments contained in this solution are sufficient for ZooMS. We refer to this method as the "AmBic" protocol.

Other non-invasive methods for ZooMS analysis have been more recently reported in which collagen was extracted using triboelectric charge generated by polyvinyl chloride erasers [5] or was retrieved directly from plastic sample storage bags [28]. These methods are very useful for specific types of samples (e.g. parchment) or in specific circumstances (when even soaking in AmBic is not possible), but have not yet been tested or optimized to be applicable to large numbers of bones.

Finally, a new protocol using bead based protein purification, referred to as SP3 (single-pot solid-phase-enhance sample preparation), was introduced recently for proteomic sample preparation [29]. Its application to ancient bone samples was described by Cleland [30]. SP3 uses a mix of hydrophilic and hydrophobic beads (coated with carboxylate functional groups) that capture proteins and allow for protein purification and buffer exchange before digestion. SP3 has gained popularity as it provides non-discriminatory and efficient binding of proteins, alongside the removal of humic substances. In the present study, we report the application of an SP3 protocol to ZooMS for the first time.

The effectiveness of the main ZooMS protocols, in terms of protein

extraction completeness, contaminant removal and success rates with regards to taxonomic identification, have not been tested in a comprehensive way for a large number of samples with different preservation status. Here, we performed a comparison between the Ambic protocol, the acid-insoluble and acid-soluble HCl-based variants, and the SP3 protocol using 400 unique bone samples from seven sites across the world.

1.2. Screening methods for bone collagen preservation

While collagen is the dominant protein in modern bone, its preservation in ancient bone varies considerably based on burial time, environment, and taphonomic processes. Theoretical models have been put forward over the years to explain collagen survival rates e.g. [31], and while these may offer reasonable predictions, collagen preservation does not often follow expected theoretical patterns [32,33].

Collagen extracted from bone is widely used in archaeology, not only for ZooMS and palaeoproteomic analyses, but also in radiocarbon (^{14}C) dating and stable isotope analyses. This has led to the development of various pre-screening methods to assess bone collagen content prior to the application of lengthier, more involved and destructive specialized preparation protocols. These screening methods include, among others, the measurement of the nitrogen content of untreated bone powder (that is the ratio of nitrogen mass to sample mass, hereafter %N) [34,35], microporosity [36], FTIR and NIR spectroscopic analyses [37–40], UV-stimulated fluorescence [41] and sulfur speciation mapping [42].

Recently, FTIR was used as a screening tool for assessing bone collagen preservation prior to ZooMS [37]. However FTIR analyses have been associated with inter-instrumental variability and only provide semi-quantitative results, which require the establishment of calibration curves for each set of samples. On the other hand, Harvey et al. [43] used ZooMS as a screening tool prior to radiocarbon dating and found good agreement between samples that produced ZooMS identification and radiocarbon dates, and suggested that ZooMS can be used additionally as a collagen predictive tool for other types of analyses.

Here, we take a different approach and use %N as an indicator of collagen preservation, juxtaposing it to the success rate of the various tested ZooMS protocols. The %N method is a commonly employed technique because it is comparatively simple, rapid and cheap, and requires between ~1-2 mg of sample. It is considered a reliable proxy for assessing bone collagen presence, especially when more than one location on the bone is analyzed [34,35]. The method does not require calibration curves for collagen quantification such is the case with FTIR-based approaches. In addition, the detection and quantification limits are lower than those achieved with other methodologies, while variability across instruments is minimal with the use of appropriate standards.

Fresh modern bone contains about 3.5–4.5% nitrogen by weight corresponding to ~22% of total collagen content [44,45], most of which is contributed by bone proteins. After the death of an organism, its nitrogen content is expected to decrease with time due to protein degradation and removal. In the past, nitrogen content of a bone was used as a relative "dating" tool [46,47]. These approaches are now superseded as it is clear that nitrogen concentration, even for bones from similar environments and age, will exhibit great variability due to ambient temperature, soil pH, collagen-degrading microorganisms, and water. Previous work indicates that, during artificial degradation of modern bone, collagen quality is not affected significantly until less than 1% of collagen remains (~0.76%N) [48]. In practical terms, %N values do not indicate collagen quality but quantity, hence we hypothesize that the higher the %N value of a bone, the higher its chances to be analyzed successfully using ZooMS. If this was the case, %N testing could be used as a cheap screening tool prior to ZooMS analyses.

2. Materials and methods

2.1. Materials

The bones included in this study come from seven archaeological sites around the world (Table 1). Their age spans from ~2000 to as old as 500,000 years ago. The majority of sites are located in parts of the world (Fig. 1), including tropical and savanna environments, in which collagen preservation is challenging and not always assured. In order to combat this and provide an effective means of comparing all four protocols, samples from Denisova Cave, a site from where multiple ZooMS projects have been successful [10,49], were included. All bone samples analyzed as part of this project were fragmentary, measuring between 1 and 4 cm in length and lacked any diagnostic morphological features. We excluded obviously burnt bones, but no other criteria were used for the selection or exclusion of samples in the study.

In addition to the "unknown" archaeological samples, a bone of *Ovis aries* from an archaeological site in Mongolia was also used. This bone is used routinely in the MPI-SHH Palaeoproteomics laboratory as a standard (positive control) to evaluate the efficacy and reproducibility of different extraction methodologies.

2.2. Reagents and software

Trifluoroacitic acid (TFA), ammonium bicarbonate (Ambic), ethanol and ethylenediaminetetraacetic acid (EDTA) were purchased from Fisher Scientific (Germany). Acetonitrile (ACN) and methanol were purchased from Carl Roth (Germany). Hydrochloric acid (HCl), C18 ZipTip pipette tips, tris(2-carboxyethyl)phosphine (TCEP), chloroacetamide (CAA), and MS grade trypsin were purchased from VWR International. α-cyano-4-hydroxycinnamic acid (CHCA) and Guandine hydrochloride (GuHCl) were purchased from Sigma-Aldrich (Germany). Molecular Weight Cut Off (MWCO) (30 kDa) ultrafiltration units were purchased from Sartorius Stedim (Germany). Carboxylate-modified paramagnetic beads (Sera-Mag Speed- Beads (Hydrophilic), CAT# 45152105050250, and Sera-Mag SpeedBeads (Hydrophobic), CAT# 65152105050250) were purchased from GE Healthcare (Germany). C18 Attract SPE Disks Bio were purchased from Affinisep.

The software used in this study includes mMass (V.5.5.0) [59], Byonic Viewer (V3.4-55-g551bce936c), Geneious Prime (V.2019.2.3), Bruker Daltonics flexAnalysis (Version3.4), and in-house ZooMS identification software. Analysis was conducted in R [60] and figures were produced using the package ggplot2 [61].

2.3. Measuring %N

The values of %N of all 400 bones included in this study were measured following the protocol of Brock et al. [34]. The bone surface was initially cleaned by air abrasion using inert aluminum oxide powder. Bone powder was drilled from a single point of each sample using a tungsten carbide spherical burr drill bit. Around 2.0 mg of bone powder was weighed and packed into pre-cleaned tin capsules. The nitrogen content of the bone powder was measured using an automated carbon and nitrogen elemental analyzer (Carlo Erba EA1108) at the Research Laboratory for Archaeology and the History of Art, School of Archaeology, University of Oxford. An in-house alanine standard (Merck, 05129, UK) was used for instrument calibration and quality control.

2.4. Experimental design for ZooMS protocols

Our experimental set up included 3 Phases which involved the application of four protocols (Fig. 2).

One to three bone chips (15–25 mg each) were extracted from each bone fragment. The number of sub-samples extracted from the same bone depended on both the overall number of protocols every single sample underwent, as well as the size limitation of the bone fragments.

Table 1
List of sites and details of material included in this study.

Site Name	Site Code	Location	Context and age of sampled context	Environment	Sample number included in the study	References
Denisova Cave	DC	Russia	Layer 14, East Chamber, ~190 ka BP	Forest-steppe, Sub-alpine	198	Douka et al. 2019 [49]; Jacobs et al. 2019 [50]
Wonderwerk Cave	WW	South Africa	Excavation 1, Stratum 6, Depth 15–20, Sq. P34, 500–230 ka BP Excavation 1, Stratum 7b, Depth 0-5 cm, Sq. T31, 500–230 ka BP Excavation 2, Stratum 2, Depth 0-5 cm, Sq. Y49, ~240–150 ka BP Stratum 2, Depth 9-10 cm, Sq. X48, 240–150 ka BP	Savanna	100	Horwitz and Chazan 2015 [51]; Chazan et al. 2020 [52]
Riparo Mochi	MOCH	Italy	Unit G9, 46–42 ka BP Unit H2, 44–41 ka BP	Temperate northern Mediterranean zone	37	Douka et al. 2012 [53]
Fa-Hien Lena	BYP	Sri Lanka	Context 253, Square P3, 48–45 ka BP Context 116, Square P6-Q6, 6 ka BP	Tropical rainforest	36	Wedage et al. 2019 [54]
Pamwak	PWK	Manus, Papua New Guinea	~26–5 ka BP	Tropical lowland	19	Fredericksen et al. 1993 [55]
Matenkupkum	MKK	New Ireland Papua New Guinea	Layer 1–4, Sq. G ~12–10 ka BP	Coastal lowland	6	Gosden and Robertson 1991 [56]
Matenbek	MBEK	New Ireland Papua New Guinea	Layers 1–2, SQ. B ~6–2 ka BP	Coastal lowland	4	Allen et al. 1989 [57]; Allen 1996 [58]

Fig. 1. World map with location of archaeological sites where the samples in this work were excavated from. For details, see Table 1 and cited references.

Fig. 2 shows the sample number per site per protocol. For every 24 samples, one blank was included in all phases. In addition, the *Ovis aries* standard was ground to powder and weighed to three weight groups: 4–5 mg, 8–10 mg, 17–19 mg; each group has four replicate aliquots. Hence, 12 standards of this bone underwent extraction with all four experimental protocols as a positive control test.

Phase 1 involved the use of the ammonium bicarbonate (AmBic) protocol. All 400 samples were analyzed in this phase. This protocol is the least destructive, least expensive, and the quickest protocol.

Phase 2 was designed to assess the efficacy of acid-based methods on

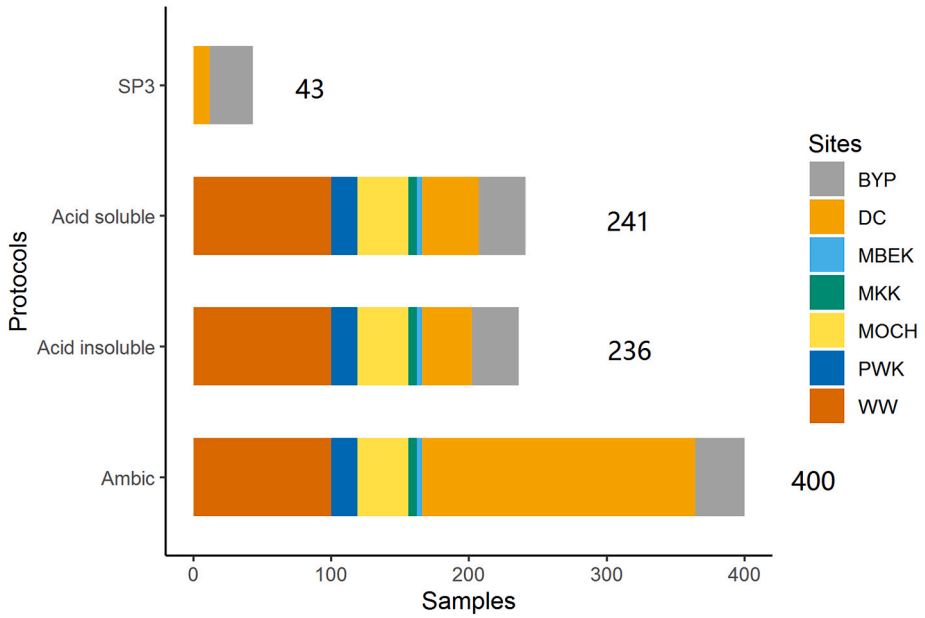


Fig. 2. Sample numbers per site for the four protocols we used in this study. The sites include: Denisova Cave (DC), Wonderwerk Cave (WW), Riparo Mochi (MOCH), Fa-Hien Lena (BYP), Pamwak (PWK), Matenkupkum (MKK), Matenbek (MBEK).

samples which failed using the AmBic protocol in Phase 1 ($n = 214$). A subset of samples that worked well in Phase 1 were included for comparison ($n = 23$ for acid insoluble and $n = 27$ for acid soluble). This phase involved an acid demineralization step followed by separate analysis of the acid-insoluble fraction and acid-soluble fractions.

Phase 3 involved testing the SP3 method to assess if it could be useful for low abundant collagen samples as it is reported to be unbiased and efficient in binding trace amounts of proteins [30,62]. A subset of samples which failed in previous protocol(s) were included ($n = 31$), along with a comparative set of well-preserved samples which were successful in both Phase 1 and Phase 2 ($n = 12$).

2.5. Methods: collagen extraction protocols

2.5.1. Phase 1

AmBic protocol [63] (after van Doorn et al. [22]): Each bone sample was covered in 100 μL 50 mM ammonium bicarbonate (NH_4HCO_3) at room temperature overnight, to clean and remove soluble contamination. The supernatant was discarded and an additional 100 μL of 50 mM NH_4HCO_3 was added in. Following incubation at 65 °C for 1 h, the supernatant was collected and was digested with 0.4 μg trypsin at 37 °C for 18 h. 1 μL 5% trifluoroacetic acid (TFA) was added to end the digestion. The digested samples were concentrated and desalted using C18 ZipTips per manufacture instruction, and eluted in 50 μL 50% ACN/0.1% TFA (*v/v*). The AmBic protocol can be accessed online at dx.doi.org/10.17504/protocols.io.bf5djq26.

2.5.2. Phase 2

Samples were demineralized in 500 μL 500 mM HCl for 1–5 days at 4 °C until the bone chips became spongy and stopped reacting. The supernatant was removed.

Acid-insoluble protocol [64] (after Welker et al. [24]): After demineralization the samples were rinsed 3 times using 50 mM NH_4HCO_3. The samples were incubated at 65 °C for 1 h, in 100 μL of 50 mM NH_4HCO_3 followed by trypsin digestion and peptide purification as in AmBic protocol.

Acid-soluble protocol [65] (after van der Sluis et al. [25]): 500 μL of the collected HCl supernatant, was transferred to a 30 kDa ultrafilter. Centrifugation allowed the removal of the HCl and any low weight molecules and facilitated buffer exchange, a total volume of 750 μL 50 mM NH_4HCO_3 was used to rinse each sample 3 times. 100 μL NH_4HCO_3 was added to the high molecular fraction remaining in the ultrafilter followed by trypsin digestion and peptide purification as in AmBic protocol. Both acid based protocols can be accessed online at dx.doi.org/10.17504/protocols.io.bf5djq26.

2.5.3. Phase 3

Single-pot solid-phase-enhanced preparation [66] (after Cleland et al. [30] and Hughes et al. [29,67]): Samples were demineralized in 500 μL 500 mM ethylenediaminetetraacetic acid (EDTA) for 5 days, on an end-over-end rotator at room temperature. 100 μL of the supernatant was decanted and the remaining supernatant and pellet were mixed with 200 μL 6 M guanidine HCl for protein solubilizaton. Prior to binding to the magnetic beads, any proteins were reduced and alkylated for 10 min at 99 °C in 10 mM Tris(2-carboxyethyl)phosphine (TCEP) and 10 mM chloroacetamide (CAA). Hydrophilic and hydrophobic paramagnetic beads were used in a ratio of 1:1 (*v/v*). 20 μL of 20 μg/μL bead mix was added to each sample. Following this, 350 μL 100% ethanol (EtOH) (*v/v*) was added to reach a final concentration of 50% ethanol in the tube, proteins were allowed to bind with the beads for 5 min in a Thermoshaker at room temperature. The sample-beads mixture was placed on the magnetic rack for 2 min to collect the beads and the supernatant was discarded. The residue was rinsed 3 times with 200 μL 80% EtOH (*v/v*). The beads were resuspended in 75 μL of 100 mM NH_4HCO_3 and digested with 0.4 μg trypsin for 18 h at 37 °C. Samples were acidified using 30 μL of 5% TFA and placed on a magnetic rack to collect supernatant for C18 peptide purification. The *Ovis aries* test samples were purified using ZipTips as described in AmBic protocol. The archaeological samples were purified using StageTips as described in the SP3 standard protocol.

The StageTips were made from two trimmed Attract SPE Disks Bio, 25 mm C18 disks in 200 μL tips. They were activated using 150 μL methanol (MeOH), 150 μL 50% ACN/0.1% TFA (v/v) and 150 μL 3% ACN/0.1% TFA (v/v) and were centrifuged at 2000 rpm for 1 min separately. Samples was then loaded onto the StageTips, washed twice with 150 μL 3% ACN/0.1% TFA (v/v) and eluted using 50 μL 50% ACN/0.1% TFA (v/v) at 2000 rpm for 5 min, 3 min and 2 min each. The SP3 protocol for ZooMS used here can be accessed online at dx.doi.org/10.17504/protocols.io.bf6pjrdn.

2.6. MALDI-TOF

All tryptic digested extracts were diluted 10× with 50% ACN/0.1% TFA (*v/v*), and mixed with an equal volume of α -cyano-4-hydroxycinnamic acid solution (10 mg/mL in 50% ACN/0.1% TFA (*v/v*)). 1.5 μL of the mixture was spotted on a Bruker ground steel plate in triplicate. Samples were measured using an Autoflex Speed LRF Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometer (Bruker). The resulting mass spectra were analyzed using mMass (V.5.5.0) [59] together with Byonic Viewer (V3.4-55-g551bce936c).

2.7. Peptide mass spectra data analysis

Samples were peak picked with a signal to noise ratio of 3.5 using mMass (V.5.5.0) after baseline correction and smoothing with the default parameters (Baseline correction precision 100 and relative offset 0. Smoothing method Savitzky-Golay algorithm, window size 0.2 *m/z* and 1.5 cycles. Peak picking with deisotoping tool, relative intensity threshold 0.5% and picking height 80%). Technical spotting replicates were averaged to get a value for each sample. Four different quality assessment criteria were used.

(1) Total number of peaks in the spectrum.
(2) Total number of ZooMS COL1 diagnostic markers (9 markers maximum, 6 peptides have individual masses and 3 peptides each have 2 masses at 16 Da difference resulting from variations in the number of oxidized prolines on the peptide [68]). These were determined by comparing the peak lists to the published diagnostic mammal markers [1,7,13,43,69–72] with a search tolerance of ±0.3 m/z.
(3) The number of non-diagnostic COL1 peaks. These were determined by choosing ten spectra from Denisova Cave, removing the contaminant and diagnostic peaks and comparing the remaining peaks. LC-MS/MS spectra of bone collagen from two bovine species (*Ovis aries* and *Connochaetes taurinus*) were analyzed in Byonic using a database of published collagen sequences (*Bos taurus* (P02453, P02465) and *Ovis aries* (XP_011983013.1, XP_004007775.1)) based on the following parameters: precursor mass tolerance 10 ppm, fragment tolerance 0.05 Da, fully specific tryptic digestion with 2 missed cleavages, 1 rare post translational modification (PTM) (rare: Glu/Gln -> pyo-Glu at N-Terminus, acetyl at the N-Terminus), 5 common PTMs (oxidation of methionine and proline, deamidation of asparagine and glutamine). The results were screened to identify collagen tryptic peptides with masses matching the shared MALDI peaks. Any masses that did not match peptides in the LC-MS/MS data were removed from the list. The sequences of these candidate shared collagen peptides were compared against the published collagen sequences of a broad range of mammals (*Bos taurus* (P02453, P02465), *Capra hircus* (XP_017920382.1, XP_005678993.1), *Ovis aries* (XP_011983013.1, XP_004007775.1), *Odocoileus virginianus* (XP_020769668.1, XP_020769440.1), *Acinonyx jubatus* (XP_026 889851.1, XP_026927691.1), *Canis lupus* (Q9XSJ7, O46392), *Equus caballus* (XP_023508478.1, XP_001492989.1), *Homo sapiens* (P02452, P08123), *Sus scrofa* (BAX02568.1, BAX02569.1)). Only masses that corresponded to peptide sequences shared by all species were then used as non-diagnostic COL1 peaks (Table S1). After the list was generated, the peak lists for all spectra were compared to the list of non-marker COL1 peaks with a search tolerance of ±0.3 *m/z*. The data from *Connochaetes taurinus* can be found at PXD020810 through MassIVE (doi:10.25345/C5TJ3M password: reviewer), sample OM 20A, DA383 and from *Ovis aries* at PXD020809 through MassIVE (doi:10.25345/C5Z75H password: reviewer), sample DA157 and DA327.
(4) The number of contaminant peaks (including peaks relating to the matrix and the containment or added proteins: trypsin, human keratins, bovine serum albumin, and bovine casein αS1) (Supplement Note 1). We compared the peaklist for each sample against the common contaminants (Big-Pig-Mix) in mMass (V.5.5.0) [59] which derives from the contaminants in Keller et al. [73] with a search tolerance of ±0.3 m/z.

3. Results

3.1. Nitrogen percentage

All samples were analyzed for their %N ($n = 400$), ie the bone mass to nitrogen mass ratio. Their values ranged between 0.02 and 3.87%N. This very broad range is indicative of the variable preservation of bone collagen from the analyzed sites. The error associated with these measurements might be up to 0.1% although the repeatability of a measurement consists of three main parameters. These include the instrumental measurement error which is normally small, ca.0.05% based on gas pulses; inhomogeneity in the bone itself due to differential preservation and particles size variation of the drilled sample; but most importantly, weighing errors for samples and standards, which will cause the largest source of error since the %N calculates a percentage by weight. Water content, in particular, can vary significantly on the day of weighing since bone powder samples may readily absorb atmospheric humidity. The alanine standards we run alongside the archaeological samples gave a mean %N value of 15.72 with a standard deviation of 0.30 ($n = 127$). Most of the variance comes from relatively few, but rather deviant, measurements; the majority of the alanine N values cluster quite tightly. Whether this variance was caused by weighing or combustion issues we have no evidence either way.

Large differences in the N values were observed within a single site, which may be linked to the different preservation conditions between stratigraphic levels, the treatment of the bone before burial and the local burial conditions (Fig. 3).

Denisova Cave (DC) has the largest sample size ($n = 198$) with all samples from the Middle Pleistocene layer 14 of the site's East Chamber (Table S2). The %N values of the DC samples range between 0.04 and 3.87%. Despite coming from the same archaeological layer, they fail the Shapiro-Wilk normality test ($W = 0.89129$, $p < 0.05$), which confirms that the data are not normally distributed, verifying the heterogeneity of collagen content even in bones from the same archaeological layer. Recent radiocarbon and genetic age estimations, optically stimulated luminescence dating, and micromorphological studies have shown that while occasional mixing between stratigraphic layers has occurred, these instances are limited [49,50,74]. Rather, local micro-environmental context as well as pre-burial conditions are likely to have played a more significant role in the variation of %N samples from Denisova Cave.

Some samples from Wonderwerk Cave (WW) demonstrated exceptionally high %N values despite the dry environment of the locality and the old age of the samples (Table S2). The %N from Stratum 2 (squares X-48 and Y-49, WW_XY) ranged between 0.10 and 3.17% (mean = 2.07), which is significantly higher than the values for samples from much older Strata 7 and 6b (squares P-34 and T-31, WW_PT) that ranged between 0.03 and 0.14% (mean = 0.08, after the removal of a single outlier which had a value of 2.41).

Most bones from Fa-Hien Lena (BYP) yielded N values below 1% with a mean value of 0.56%, suggesting that this site has moderate collagen preservation. Four samples from BYP yielded values above 1% (Table S2). All samples from remaining sites, Riparo Mochi (MOCH), Pamwak (PWK), Matenkupkum (MKK) and Matenbek (MBEK), had nitrogen values less than 0.1%, indicating significant loss of collagen (Table S2).

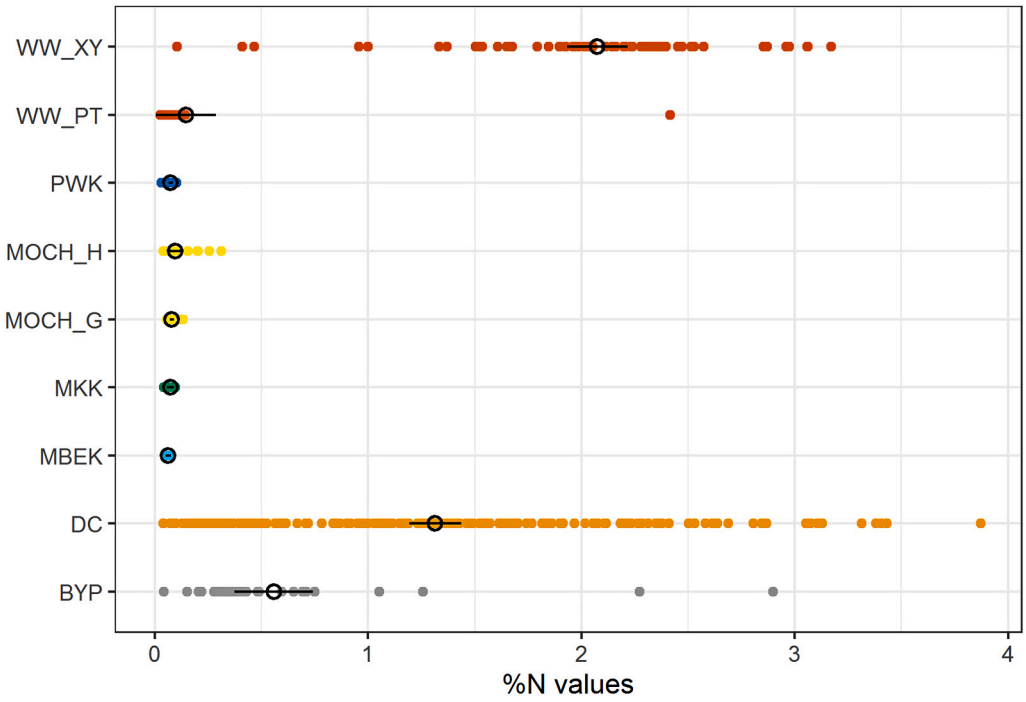


Fig. 3. Nitrogen percentage (%N) values for all samples included in this study separated by site. The sites include: Denisova Cave (DC), Wonderwerk Cave (WW), Riparo Mochi (MOCH), Fa-Hien Lena (BYP), Pamwak (PWK), Matenkupkum (MKK), Matenbek (MBEK). Open, black circles indicate mean values, the black line indicates 95% confidence intervals. Wonderwerk Cave (WW) and Riparo Mochi (MOCH) are separated based on their archaeological context.

3.2. ZooMS results

After confirmation from the LC-MS/MS data, 21 non-diagnostic collagen peaks were identified and then used to assess spectral quality (Table S1). Multiple blanks were analyzed alongside all samples, and all returned negative results, hence no cross-contamination was observed. The *Ovis aries* standard yielded positive results from all protocols.

Although performing taxonomic identifications is not the aim of the current study, it is the primary determination of ZooMS being successful. Therefore we used both the ability to ID the spectra and the presence of ZooMS diagnostic peptide markers as the criteria to assess each protocol's efficacy. Identification was attempted for all samples (Table S3). If the samples could be assigned to any taxonomic level below mammal in the ZooMS reference library the samples were considered IDable, otherwise they were regarded as a Failed. Since database bias could result in high quality spectra being classified as Failed using this method, we further assessed the results of this metric for all samples in two ways. First, we analyzed samples that failed for the number of peptide markers and shared collagen peaks. Only two samples (BYP24, DC5860) had both shared and common collagen peaks indicating that they likely derive from a taxon not present in the database. Across all methods, only 16 other samples had marker peaks identified but were classified as Failed. Of these the majority ($n = 10$) were IDed subsequently using another protocol. Second, we assessed how the classifications of Failed and IDable mapped onto the total number of peaks. There is a high correlation between IDable and Failed samples in relation to the total number of peaks present(Fig. 4, S2). While we cannot exclude the possibility that some of the samples derive from non-mammal species, there is no indication for the samples used in this study that the classification of IDable and Failed was driven by database bias. Therefore, as IDable and Failed are essential classifications for ZooMS research, and rely on database availability, subsequent results are discussed primarily in relation to these categories.

In Phase 1, all 400 samples were extracted using the Ambic protocol, and 219 failed. The IDable 181 samples were all from DC. Of these samples, 147 had 8 or more diagnostic markers present and the remaining 34 had 4–7 diagnostic markers present (Fig. S1).

In Phase 2, we assessed samples that failed AmBic protocol ($n = 214$) and selected a small number of IDable samples from DC (acid-insoluble protocol: $n = 23$, acid-soluble protocol: $n = 27$). We used two protocols that involve a demineralization step prior to further collagen pretreatment. Of the IDable spectra from Phase 1, the majority of the samples ($n = 26$) were IDable the acid based methods that were completed ($n = 19$ with both acid methods, $n = 7$ with 1 method). Of the samples that failed the AmBic protocol, 14 samples (11 from BYP and 3 for DC) were IDable in one or both of the acid based methods.

The acid-insoluble method yielded slightly better results over the acid-soluble one. Of the 14 additional passing samples in Phase 2, all yielded passing spectra from acid-insoluble protocol, while only 7 yielded passing spectra from acid-soluble protocol. The remaining 200 failed samples from Phase 1 also failed in both protocols of Phase 2. Considering that we expected the samples from MOCH, PWK, MKK, and MBEK, and one group of samples from WW to exhibit low collagen preservation due to their very low %N, these negative results while disappointing are not entirely surprising.

In order to test specifically to see if SP3 performed better on samples with low collagen preservation, for Phase 3, we analyzed 12 samples from DC with good collagen preservation and 31 samples from BYP which had medium-to-low collagen preservation (mean $\%N = 0.56$ for the site; Fig. 3). All 12 samples from DC generated IDable spectra; however, several of the low mass markers were missing or have much lower intensities compared with the spectra in previous phases similar to what was observed in the control sample. Of the BYP samples, only one (BYP06) yielded IDable spectra in Phase 3, this sample was also IDable in both acid fraction in Phases 2. BYP14 was IDable in Phase 2, but in Phase 3 it gelatinized upon the addition of ethanol, possibly as a result of bead overloading or co-extracted compounds and needed to be excluded. The remaining BYP samples all failed.

The overall success rates for the sites including all methods are ~90% for DC, ~30% for BYP, and 0% for all of the other sites tested (Table S4).

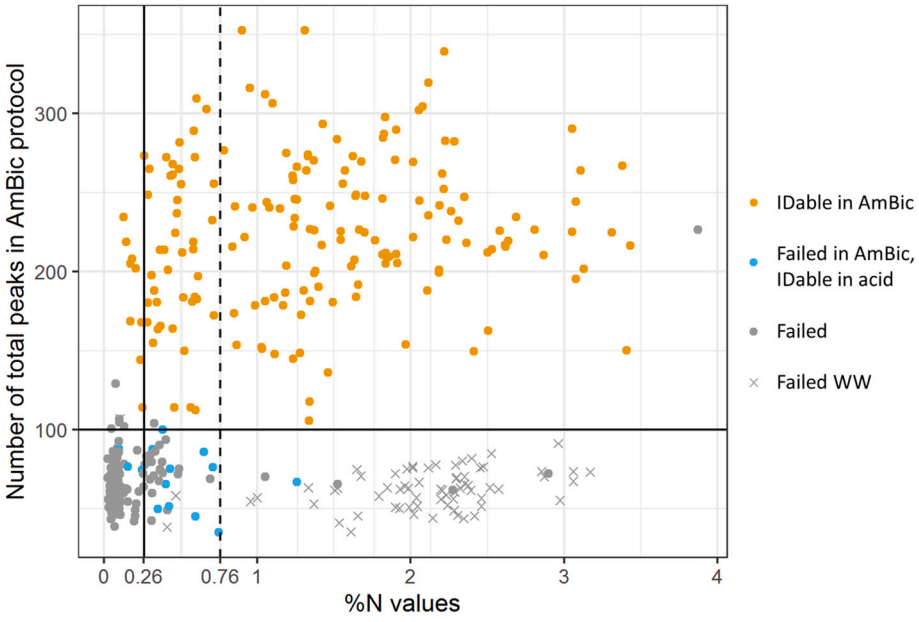


Fig. 4. The nitrogen threshold for the AmBic protocol for 400 samples from seven sites. The orange circles relate to the 181 samples that passed Ambic protocol, the blue circles correspond to samples that failed AmBic protocol but were IDable for acid-based protocols (11 from BYP and 3 from DC). These samples yielded higher numbers of total peptide peaks in acid-based protocols. The grey circles correspond to samples that failed all tested protocols. The grey X marks correspond to samples with elevated %N values from WW_XY for which we suspect nitrogen contamination. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Correlation of %N with number of peptide peaks

Given that the number of ZooMS markers or total number of peptides alone cannot quantify the amount of collagen retrieved from a sample, we compare the number of total peptide peaks against the %N value. The number of total peaks may be considered a useful proxy for the quality of the collagen remaining in the bone and for whether a sample would be eventually IDable or not.

We correlate the %N values for all 400 samples analyzed using AmBic protocol, with the number of total peaks per sample (Fig. 4, Fig. S2). We observe that when the number of total peaks exceeds ~100, the sample is very likely to yield a successful ZooMS identification (Fig. 4, horizontal threshold). Further work is required to assess what causes this threshold and to identify the specific differences in the quality of extracted collagen.

Fig. 4 also highlights the 68 bones from WW (67 from Squares X-Y and 1 from Square P) which show abnormally high %N values yet resulted in no extractable collagen. In addition, four samples from BYP, located in a rainforest setting, also showed a similar pattern where the %N values were above 1.0, but ZooMS for both AmBic and acid protocols failed to produce IDable spectra. We suspect that the exceptionally elevated %N values in these samples are due to environmental contamination. This may be linked to exposure of bones to bird or bat guano-rich sediment that contributes high levels of nitrate and ammonium contamination and can alter significantly and sometimes irreversibly change the original nitrogen concentration of a bone. Indeed, the archaeological layers of WW were exposed for the first time during illicit guano digging operations when the site's sediment was exploited as fertilizer [51], while micromorphology has demonstrated layers with evidence for phosphatization, which may be related to decomposition of bat or bird guano [75]. It is likely that the effect of this soil contaminant is reflected in the exceptionally high %N values in our study.

Therefore, for assessing the %N threshold the samples from WW ($n = 100$) were all removed and the remaining 300 samples were divided into "IDable" and "Failed" groups, 95% of the IDable samples fall above 0.26%N in the AmBic method. To explore this threshold, we evaluated the data for true positives (percent of the total IDable samples that were above the chosen %N value) and the true negatives (percent of the total failed samples that were below the chosen %N value) between 0.10%N (the lowest %N value of an IDable spectra) and 0.76%N (where the true positive rate dropped to 70%) for the AmBic protocol and combining the results from all protocols (Fig. S4). For the AmBic protocol, 100% of the IDable spectra have a %N value greater than 0.13, 95% greater than 0.26%N, 80% greater than 0.51%N, and 70% greater than 0.76%N (Fig. 4, vertical thresholds at 0.26 and 0.76%N). Of the 300 randomly chosen samples across seven sites using the AmBic protocol, with no thresholding, 60% of the total samples analyzed were IDable. Using the threshold at 0.26%N where only 5% of the possible IDable spectra would be excluded from analysis, the percent success rate of the total samples analyzed increased to 82% for just the AmBic protocol and 88% when including the results from all protocols. The threshold can be lowered to include 100% of the samples that would be IDable in any method at the expense of a lower overall success rate (75%) and increased to get success rates of ~90% at the expense of the exclusion of ~20% of the IDable spectra. Our results indicate that screening for collagen preservation before ZooMS can drastically increase the success rate (Fig. S3). For large ZooMS sample sets this is a huge improvement in time and cost expended.

There is a certain degree of variability in the collagen content and, by extension, in the %N values across a bone [35]. This is probably a result of both bone formation and diagenetic effects. The internal variability means that %N values cannot always be used at face value for an entire bone, but in the case of ZooMS if microsampling is performed very close to the area analyzed for %N the generated result should correspond to the prediction here.

We caution that the value of 0.26%N is based primarily on the results of a single site, Denisova Cave, which enjoys a high level of collagen preservation. The majority of sites included in this study were located in environments where collagen preservation is known to be poor and more work is therefore required to establish whether the 0.26%N value is site-dependent (i.e. inherently linked to conditions impacting local collagen preservation), or is affected by inter-user variability too, since accurate sample weighing will have a significant effect in the accuracy of sample mass to nitrogen mass ratio.

In the radiocarbon community, samples with more than 0.76%N are considered likely to retain more than 1% collagen, which is typically required for successful ^{14}C analysis [34]. For successful ZooMS analysis we determined that a lower value (more than 0.26%N) is all that is needed. In fact, only ~70% of the IDable spectra had %N values greater than 0.76. Our findings therefore do not support previous claims [43] that successful ZooMS identification could be a predictor of successful

radiocarbon analyses. While it is highly likely that a failed ZooMS result predicts a failed radiocarbon result, the reverse is not true. For well-preserved samples with %N values above 0.8%N both ZooMS and radiocarbon will likely work; however, in the intermediate range between ~0.2–0.8%N, bones that produce ZooMS, will likely fail to yield enough collagen for other types of collagen-based analyses. Screening methods such as %N, which we use here, and FTIR [76] are much better suited to reliably identify samples for analyses (ZooMS, ^{14}C, stable isotopes).

4.2. Assessment of the extraction protocols

The *Ovis aries* standard yielded all 9 diagnostic markers (9 peptides with 12 masses) for three weight categories (4–5 mg, 8–10 mg, 17–19 mg) for the Ambic and both acid-based protocols (Fig. S5-S8, Table S5) with similar patterns of intensities. For SP3, the highest weight category had 6 diagnostic markers (6 peptides with 8 masses). The lower two weight categories had 7 and 6 markers each and the intensities of markers were also lower than for the other protocols. Generally the starting weight of mammal bones for ZooMS is between 10 and 30 mg of material and so would fall into the largest weight category. Our results indicate that for well-preserved samples, smaller starting weights are possible and indeed more favorable to avoid overloading the ZipTips. SP3 is generally considered to work best at very low protein concentrations, our results show that the lower weight categories do have the same markers but lower overall intensities than the highest weight category.

When assessing the quality of the SP3 data, we analyzed the standard as well as the 12 well preserved samples from DC. A drastic decrease in the amount of peaks in a key mass region (m/z 1170–1260) were detected in all SP3 samples. This area contains the COL1α2 978–990 (A) diagnostic markers which often separate large taxonomic groups (Fig. S9). In addition, the intensities of the diagnostic markers for SP3 showed a different overall pattern than for the three other protocols.

To assess and compare the different collagen extraction protocols we chose a sample group consisting of 13 samples from BYP, and 24 samples from DC. These 37 samples (%N: 0.1–3.4, mean = 0.91) were chosen because they had been extracted with at least three protocols each and generated marker peaks with at least one protocol. For each method, all four quality assessment criteria were determined: the total number of peaks, number of non-diagnostic COL1 peaks, number of contaminant peaks, and number of diagnostic markers (Table S6, Fig. 5).

For the archaeological samples, the commonly used ZooMS methods yielded similar results across the four quality assessment criteria we employed here (Fig. 5A-B). Four runs of one-way ANOVA were processed to analyze whether the overall effect of "protocol" is significant. The F-statistic values of total peak, non-marker collagen peaks, contamination peaks and marker peaks are from 7.93 to 28.45, with all of the p-values<0.05, indicating there were significant differences between protocols in four different quality assessment criteria. Planned pairwise comparison t-tests (Bonferroni test) demonstrated that there was no significant difference in the number of peaks between the AmBic, acid soluble, and acid insoluble protocols by any metric (t-statistic values from 0.18 to 2.37, p-values>0.12). SP3 had significantly fewer peaks compared with AmBic and acid insoluble protocols by all metrics and significantly fewer peaks compared with the acid soluble protocol for the total, non-marker, and contaminant peaks (t-statistic values from 4.19 to 8.29, p-values<0.05). For the marker peaks SP3 showed no significant difference from the acid soluble protocol, the t-statistic is 2.21 and p-values = 0.18. This is not surprising because there are several samples extracted with acid-soluble yielded only few marker peaks.

In this study, most analyzed samples that were IDable were obtained using the non-destructive, generally less expensive and time consuming, AmBic method. The addition of the demineralization step in Phase 2 increased the number of IDable samples by 14, all except one of which had values below 0.76%N (Fig. 4). Although the sample size is non-random and small, the acid insoluble protocol (14/14 additional IDable samples) outperformed the acid soluble protocol (7/14 additional IDable samples). Both acid-based approaches use demineralization in HCl and yield similar peptide mass fingerprints (e.g. Fig. 6 and Fig. S10). Although more work is necessary to better understand the conditions and quality of samples for which a demineralization step will improve ZooMS spectra, our work supports the conclusion that the acid-based protocols can improve identification for samples with poor collagen preservation. Buckley et al. [72] suggested that the acid insoluble approach is likely to reduce various forms of contamination; however, our results do not support this as both methods show similar amounts of contaminant peptides (Fig. 5b).

The SP3 protocol underperformed in all four quality assessment criteria. SP3 was developed to use beads known for their binding ability to a wide range of proteins in cell lysates [29]. It has been shown to perform better than other protein extraction methods for low quantities of starting material [62]. However, it has been developed and tested for performance in complex mixtures with an emphasis on identifying the

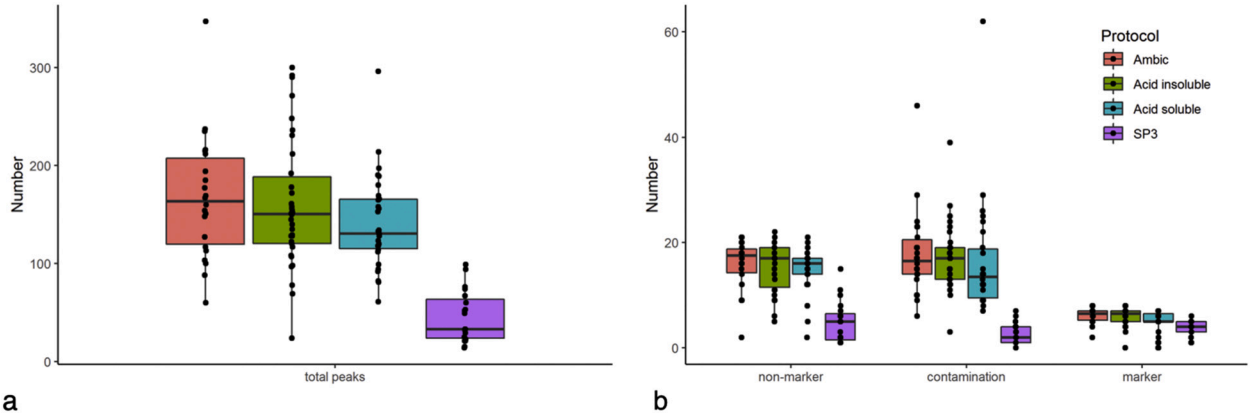


Fig. 5. Boxplot depicting the number of peaks for the four criteria we set up to evaluate each protocol. Each black point represents a sample. 37 samples (35 were IDable) are represented out of the 66 samples that were run in at least three of the three methods. The other 29 samples have no marker peaks in any protocol and therefore were excluded for the comparison. (A) Total number of peaks for each protocol. (B) Non-marker collagen peaks (maximum 21), contamination peaks, and diagnostic markers (maximum 9) for each protocol. In each boxplot, the 1st and 3d quartiles (the 25th and 75th percentiles) are shown, the whiskers represent the highest and lowermost observations, with some outliers falling out of them, and the vertical line inside the box is the median. SP3 has a significantly lower number of total peaks, non-marker peaks, and contamination peaks from the other three methods and a significantly lower number of markers than the AmBic and acid insoluble methods.

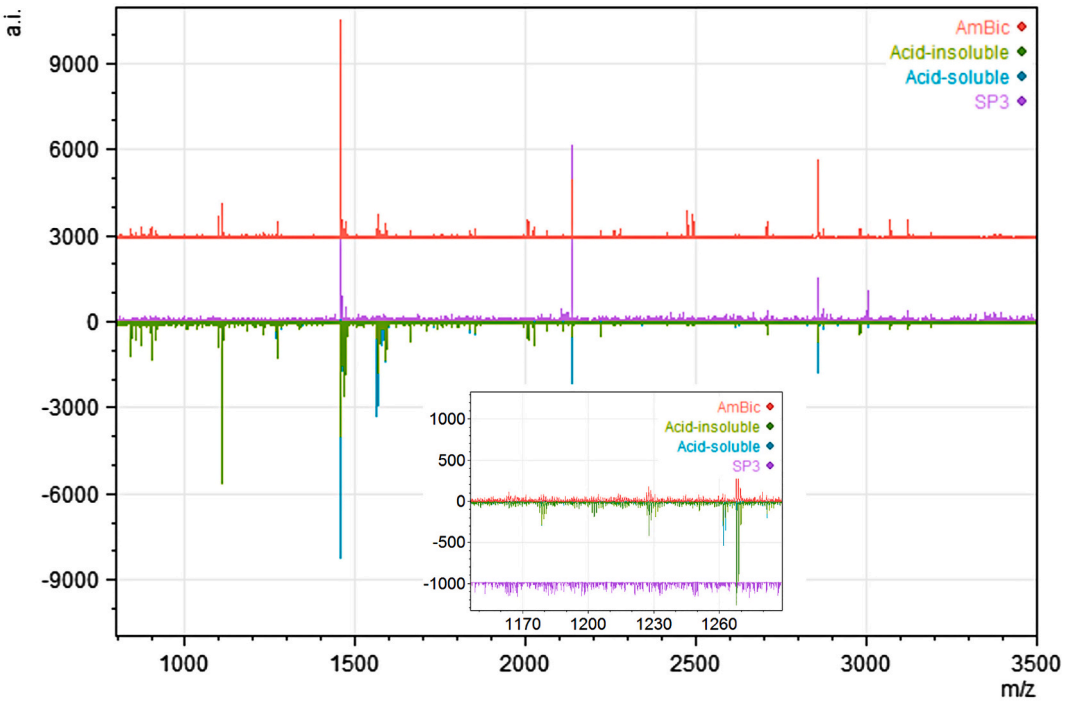


Fig. 6. Comparison of the MALDI spectra for four protocols applied on the same sample (DC5770). The Ambic protocol (orange), the acid-based protocols (green and light blue) generated 24, 24 and 20 peptides, respectively. The SP3 protocol (purple) produced 5 peptides. The insert shows the lack of peptides with masses between 1170 and 1260 using SP3 protocol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

least abundant proteins in the mixtures which are analyzed by LC-MS/MS, therefore it may be less suited to recovering the most abundant protein in archaeological bone, type I collagen for analysis on a MALDI-TOF-MS. We hypothesize that the mechanism affecting this may be both the demineralization process and the selection of specific soluble fractions by the beads. As a demineralization agent, EDTA chelates calcium removing it from the mineral lattice and releases the phosphate in solution hence resulting in the matrix demineralization. Mineral-associated proteins, instead of collagen, would benefit from EDTA-induced demineralization resulting in higher recovery rates for non-collagen proteins [30,77].

Only one of the samples that failed in AmBic was successful in SP3 (BYP20). Even in this case the number of collagen peptides was not improved over the acid based methods (23 in acid-insoluble 19 in acid soluble, and 18 in SP3). This loss of collagen peptides is also most concentrated in the m/z 1170–1260 range (Fig. S10). Additional development including LC-MS/MS analysis will be required to understand the biases in SP3 against collagen peptides.

Overall, the non-predictable reduction of total peaks, most importantly of the collagen peptide peaks and the lack of IDable spectra from low-to-medium collagen preservation samples using SP3 is worrisome and difficult to explain without further testing and peptide sequencing. Although this is so far the only application of SP3 for ZooMS, we find no evidence that SP3 improves the success rate of ZooMS in comparison to the more routinely used protocols that we have also tested here.

Finally, other factors, such as time and cost efficiency are also important when choosing an extraction protocol. The acid protocols are more time consuming than the AmBic protocol. In addition, the same sample which has been extracted using the AmBic method can be successfully re-extracted with either or both acid-based methods [13], meaning if the AmBic extractions fail resampling is not required for the acid protocol(s). When analyzing museum specimens or other types of finite and precious samples, the degree to which one can apply destructive methodologies must also be considered. The minimally destructive AmBic protocol and other experimental triboelectric protocols may be more appropriate as a first step before more invasive protocols are employed, e.g. [5]. The flexibility of ZooMS also allows for a broader consideration of the entire research approach. For instance, if the same sample has previously undergone radiocarbon dating, stable isotope analysis, or more in-depth proteomic analysis, it is possible to re-use the collagen extracted from these other methods for ZooMS taxonomic identification e.g. [3].

5. Conclusions

In the present study, we attempt a comparison of bone collagen extraction and purification protocols (AmBic, acid-soluble, acid-insoluble fractions) and, for the first time, SP3, routinely used in ZooMS. The main aims were to assess which protocol performs better in terms of identification potential and, coupled to bone collagen content, to determine a threshold of %N values which could improve ZooMS success rates.

The non-destructive protocol (AmBic) works well with well-preserved bones such as those from high-altitude environments, or younger in age. The AmBic protocol simply allows the already denatured collagen to leach out of the bone matrix and into the buffer solution. In the acid-based protocols, the bone matrix demineralization process releases collagen otherwise linked to and/or inhibited by the hydroxyapatite crystals. The acid protocols often result in higher collagen yields and better spectra, although care should be taken as it is already known that longer demineralization times may impact collagen quality. For older, more degraded samples, or samples from environments likely to accelerate protein degradation (e.g. warmer environments, such as the tropics) an acid demineralization step could prove beneficial. In such cases, using the acid-based protocols is a safer option and most likely to produce informative spectra. Acid-insoluble protocol appeared to be performing marginally better.

We also explored the potential of SP3 which has shown great potential for shotgun palaeoproteomic approaches, as an alternative ZooMS protocol. On the basis of over 40 analyzed samples, SP3 underperformed in the number of overall peaks and collagen peptide markers for each sample. SP3 is known for very successfully removing co-

extracted compounds. No interference from humic acids in the MALDI-TOF-MS spectra was observed in any of the samples tested here. As humic acids can co-extract with collagen and interfere with the MALDI-TOF-MS, SP3 could be used when samples are, or suspected to be, highly contaminated. However, it is not advisable at this stage to be employed as a routine ZooMS protocol.

Finally, we explored the correlation between N content of untreated bone powder with the overall efficacy and success rate of various ZooMS protocols. We show that using a threshold of above 0.26%N improves identification success of the tested samples by 20% while excluding only a small amount (<5%) of samples successfully identifiable with both the AmBic and acid protocols. This is particularly important for high throughput identification where thousands of samples are processed. The %N test is fast, it requires sample removal of ~1-2 mg bone powder and virtually no sample preparation, hence we suggest that this should be a first choice when screening large bone assemblages, or when deciding whether a bone is suitable for ZooMS analyses.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2020.104078.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

The MALDI spectrometric data and the nitrogen values are publicly available (https://zenodo.org/deposit/3784979#). The LC-MS/MS data from *Connochaetes taurinus* can be found at PXD020810 through MassIVE (doi:10.25345/C5TJ3M password: reviewer), sample OM 20A, DA383 and from *Ovis aries* at PXD020809 through MassIVE (doi:10.25345/C5Z75H password: reviewer), sample DA157 and DA327. All of the protocols are available on protocols.io. The AmBic, acid soluble, and acid insoluble protocols can be found at dx.doi.org/10.17504/protocols.io.bf5djq26. The SP3 protocol for ZooMS used here can be found at dx.doi.org/10.17504/protocols.io.bf6pjrdn.

Acknowledgments

This work received support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, grant agreement 715069-FINDER-ERC-2016-STG, awarded to Katerina Douka. We would like to thank the Max Planck Society and the staff of the Department of Archaeology, Max Planck Institute for the Science of Human History, for facilitating this work and providing training and access to equipment, in particular, Patrick Roberts, Jana Zech and Nicole Boivin; and the Research Laboratory for Archaeology and the History of Art, School of Archaeology, University of Oxford for providing access to their stable isotope facilities.

The Wonderwerk Cave project is funded by grants from the Canadian Social Sciences and Humanities Research Council, and The Leakey Foundation.

The Denisova Cave excavation is supported by grants from the Russian Foundation for Basic Research (RFBR), project No. 18-09-40100.

References

[1] M. Buckley, M. Collins, J. Thomas-Oates, J.C. Wilson, Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 3843–3854, https://doi.org/10.1002/rcm.4316.

[2] M. Buckley, M. Collins, J. Thomas-Oates, A method of isolating the collagen (I) α2 chain carboxytelopeptide for species identification in bone fragments, Anal. Biochem. 374 (2008) 325–334.

[3] A.N. Coutu, G. Whitelaw, P. le Roux, J. Sealy, Earliest evidence for the ivory trade in southern Africa: isotopic and ZooMS analysis of seventh–tenth century ad ivory from KwaZulu-Natal, Afr. Archaeol. Rev. 33 (2016) 411–435, https://doi.org/10.1007/s10437-016-9232-0.

[4] J.A. Ebsen, K. Haase, R. Larsen, D.V.P. Sommer, L.Ø. Brandt, Identifying archaeological leather–discussing the potential of grain pattern analysis and zooarchaeology by mass spectrometry (ZooMS) through a case study involving medieval shoe parts from Denmark, J. Cult. Herit. 39 (2019) 21–31.

[5] S. Fiddyment, B. Holsinger, C. Ruzzier, A. Devine, A. Binois, U. Albarella, R. Fischer, E. Nichols, A. Curtis, E. Cheese, M.D. Teasdale, C. Checkley-Scott, S. J. Milner, K.M. Rudy, E.J. Johnson, J. Vnoucek, M. Garrison, S. McGrory, D. G. Bradley, M.J. Collins, Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15066–15071, https://doi.org/10.1073/pnas.1512264112.

[6] D.P. Kirby, M. Buckley, E. Promise, S.A. Trauger, T.R. Holdcraft, Identification of collagen-based materials in cultural heritage, Analyst 138 (2013) 4849–4858, https://doi.org/10.1039/c3an00925d.

[7] M. Buckley, S. Whitcher Kansa, S. Howard, S. Campbell, J. Thomas-Oates, M. Collins, Distinguishing between archaeological sheep and goat bones using a single collagen peptide, J. Archaeol. Sci. 37 (2010) 13–20.

[8] G. Pothier Bouchard, J. Riel-Salvatore, F. Negrino, M. Buckley, Archaeozoological, taphonomic and ZooMS insights into the Protoaurignacian faunal record from Riparo Bombrini, Quat. Int. (2020), https://doi.org/10.1016/j.quaint.2020.01.007, 0–1.

[9] V. Sinet-Mathiot, G.M. Smith, M. Romandini, A. Wilcke, M. Peresani, J.J. Hublin, F. Welker, Combining ZooMS and zooarchaeology to study Late Pleistocene hominin behaviour at Fumane (Italy), Sci. Rep. (2019), https://doi.org/10.1038/s41598-019-48706-z.

[10] S. Brown, T. Higham, V. Slon, S. Paabo, M. Meyer, K. Douka, F. Brock, D. Comeskey, N. Procopio, M. Shunkov, A. Derevianko, M. Buckley, Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis, Sci. Rep. 6 (2016) 1–8, https://doi.org/10.1038/srep23559.

[11] S. Charlton, M. Alexander, M. Collins, N. Milner, P. Mellars, T.C. O'Connell, R. E. Stevens, O.E. Craig, Finding Britain's last hunter-gatherers: a new biomolecular approach to 'unidentifiable' bone fragments utilising bone collagen, J. Archaeol. Sci. 73 (2016) 55–61, https://doi.org/10.1016/j.jas.2016.07.014.

[12] T. Devièse, I. Karavanić, D. Comeskey, C. Kubiak, P. Korlević, M. Hajdinjak, S. Radović, N. Procopio, M. Buckley, S. Pääbo, T. Higham, Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 10606–10611, https://doi.org/10.1073/pnas.1709235114.

[13] F. Welker, M. Hajdinjak, S. Talamo, K. Jaouen, M. Dannemann, F. David, M. Julien, M. Meyer, J. Kelso, I. Barnes, S. Brace, P. Kamminga, R. Fischer, B.M. Kessler, J. R. Stewart, S. Pääbo, M.J. Collins, J.J. Hublin, Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 11162–11167, https://doi.org/10.1073/pnas.1605834113.

[14] T.D. White, M.T. Black, P.A. Folkens, Human Osteology, Academic Press, 2011.

[15] J.D. Currey, The structure and mechanics of bone, J. Mater. Sci. 47 (2012) 41–54.

[16] P. Zioupos, J.D. Currey, A. Casinos, Exploring the effects of hypermineralisation in bone tissue by using an extreme biological example, Connect. Tissue Res. 41 (2000) 229–248.

[17] K. Henriksen, Type I collagen, in: M. Karsdal (Ed.), Biochem. Collagens, Laminins Elastin. Struct. Funct. Biomarkers, Academic Press, 2019, pp. 1–11.

[18] M.D. Shoulders, R.T. Raines, Collagen structure and stability, Annu. Rev. Biochem. 78 (2009) 929–958.

[19] F. Welker, Palaeoproteomics for human evolution studies, Quat. Sci. Rev. 190 (2018) 137–147, https://doi.org/10.1016/j.quascirev.2018.04.033.

[20] N. Rybczynski, J.C. Gosse, C.R. Harington, R.A. Wogelius, A.J. Hidy, M. Buckley, Mid-Pliocene warm-period deposits in the high Arctic yield insight into camel evolution, Nat. Commun. 4 (2013) 1–9.

[21] D.A. Stover, B.C. Verrelli, Comparative vertebrate evolutionary analyses of type I collagen: potential of COL1a1 gene structure and intron variation for common bone-related diseases, Mol. Biol. Evol. 28 (2011) 533–542.

[22] N.L. van Doorn, H. Hollund, M.J. Collins, A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction, Archaeol. Anthropol. Sci. 3 (2011) 281–289, https://doi.org/10.1007/s12520-011-0067-y.

[23] K.K. Richter, J. Wilson, A.K.G. Jones, M. Buckley, N. van Doorn, M.J. Collins, Fish 'n chips: ZooMS peptide mass fingerprinting in a 96 well plate format to identify fish bone fragments, J. Archaeol. Sci. 38 (2011) 1502–1510, https://doi.org/10.1016/j.jas.2011.02.014.

[24] F. Welker, M. Soressi, W. Rendu, J.J. Hublin, M. Collins, Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France, J. Archaeol. Sci. 54 (2015) 279–286, https://doi.org/10.1016/j.jas.2014.12.010.

[25] L.G. van der Sluis, H.I. Hollund, M. Buckley, P.G.B. De Louw, K.F. Rijsdijk, H. Kars, Combining histology, stable isotope analysis and ZooMS collagen fingerprinting to investigate the taphonomic history and dietary behaviour of extinct giant tortoises from the Mare aux Songes deposit on Mauritius, Palaeogeogr. Palaeoclimatol. Palaeoecol. 416 (2014) 80–91, https://doi.org/10.1016/j.palaeo.2014.06.003.

[26] S. Weiner, O. Bar-Yosef, States of preservation of bones from prehistoric sites in the Near East: a survey, J. Archaeol. Sci. 17 (1990) 187–196.

[27] N. Van Doorn, The Applications and Limitations of a Minimally Destructive Approach to Archaeological Proteomics, University of York, 2012.

[28] K. McGrath, K. Rowsell, C. Gates St-Pierre, A. Tedder, G. Foody, C. Roberts, C. Speller, M. Collins, Identifying archaeological bone via non-destructive ZooMS

and the materiality of symbolic expression: examples from Iroquoian bone points, Sci. Rep. 9 (2019) 1–10, https://doi.org/10.1038/s41598-019-47299-x.
[29] C.S. Hughes, S. Foehr, D.A. Garfield, E.E. Furlong, L.M. Steinmetz, Ultrasensitive proteome analysis using paramagnetic bead technology, Mol. Syst. Biol. (2014) 1–14.
[30] T.P. Cleland, Human bone paleoproteomics utilizing the single-pot, solid-phase-enhanced sample preparation method to maximize detected proteins and reduce Humics, J. Proteome Res. 17 (2018) 3976–3983, https://doi.org/10.1021/acs.jproteome.8b00637.
[31] M.J. Collins, C.M. Nielsen-Marsh, J. Hiller, C.I. Smith, J.P. Roberts, R.V. Prigodich, T.J. Wess, J. Csapò, A.R. Millard, G. Turner-Walker, The survival of organic matter in bone: a review, Archaeometry 44 (2002) 383–394, https://doi.org/10.1111/1475-4754.t01-1-00071.
[32] E. Cappellini, F. Welker, L. Pandolfi, J. Ramos-Madrigal, D. Samodova, P.L. Rüther, A.K. Fotakis, D. Lyon, J.V. Moreno-Mayar, M. Bukhsianidze, R. Rakownikow Jersie-Christensen, M. Mackie, A. Ginolhac, R. Ferring, M. Tappen, E. Palkopoulou, M.R. Dickinson, T.W. Stafford, Y.L. Chan, A. Götherström, S.K.S. S. Nathan, P.D. Heintzman, J.D. Kapp, I. Kirillova, Y. Moodley, J. Agusti, R. D. Kahlke, G. Kiladze, B. Martínez-Navarro, S. Liu, M. Sandoval Velasco, M.H. S. Sinding, C.D. Kelstrup, M.E. Allentoft, L. Orlando, K. Penkman, B. Shapiro, L. Rook, L. Dalén, M.T.P. Gilbert, J.V. Olsen, D. Lordkipanidze, E. Willerslev, Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny, Nature 574 (2019) 103–107, https://doi.org/10.1038/s41586-019-1555-y.
[33] B. Demarchi, S. Hall, T. Roncal-Herrero, C.L. Freeman, J. Woolley, M.K. Crisp, J. Wilson, A. Fotakis, R. Fischer, B.M. Kessler, Protein sequences bound to mineral surfaces persist into deep time, Elife 5 (2016), e17092.
[34] F. Brock, R. Wood, T.F.G. Higham, P. Ditchfield, A. Bayliss, C.B. Ramsey, Reliability of nitrogen content (%N) and carbon:nitrogen atomic ratios (C:N) as indicators of collagen preservation suitable for radiocarbon dating, Radiocarbon 54 (2012) 879–886, https://doi.org/10.1017/s0033822200047524.
[35] E. Jacob, D. Querci, M. Caparros, C. Barroso Ruiz, T. Higham, T. Devièse, Nitrogen content variation in archaeological bone and its implications for stable isotope analysis and radiocarbon dating, J. Archaeol. Sci. 93 (2018) 68–73, https://doi.org/10.1016/j.jas.2018.02.019.
[36] F. Brock, T. Higham, C.B. Ramsey, Pre-screening techniques for identification of samples suitable for radiocarbon dating of poorly preserved bones, J. Archaeol. Sci. 37 (2010) 855–865, https://doi.org/10.1016/j.jas.2009.11.015.
[37] G. Pothier Bouchard, S.M. Mentzer, J. Riel-Salvatore, J. Hodgkins, C.E. Miller, F. Negrino, R. Wogelius, M. Buckley, Portable FTIR for on-site screening of archaeological bone intended for ZooMS collagen fingerprint analysis, J. Archaeol. Sci. Rep. 26 (2019) 101862, https://doi.org/10.1016/j.jasrep.2019.05.027.
[38] M. D'Elia, G. Gianfrate, G. Quarta, L. Giotta, G. Giancane, L. Calcagnile, Evaluation of possible contamination sources in the 14C analysis of bone samples by FTIR spectroscopy, Radiocarbon 49 (2007) 201–210, https://doi.org/10.1017/S0033822200042120.
[39] M. Lebon, I. Reiche, X. Gallet, L. Bellot-Gurlet, A. Zazzo, Rapid quantification of bone collagen content by ATR-FTIR spectroscopy, Radiocarbon 58 (2016) 131–145.
[40] M. Sponheimer, C.M. Ryder, H. Fewlass, E.K. Smith, W.J. Pestle, S. Talamo, Saving old bones: a non-destructive method for bone collagen prescreening, Sci. Rep. 9 (2019) 1–7, https://doi.org/10.1038/s41598-019-50443-2.
[41] N. Hoke, J. Burger, C. Weber, N. Benecke, G. Grupe, M. Harbeck, Estimating the chance of success of archaeometric analyses of bone: UV-induced bone fluorescence compared to histological screening, Palaeogeogr. Palaeoclimatol. Palaeoecol. 310 (2011) 23–31, https://doi.org/10.1016/j.palaeo.2011.03.021.
[42] J. Anné, N.P. Edwards, F. Brigidi, P. Gueriau, V.L. Harvey, K. Geraki, L. Slimak, M. Buckley, R.A. Wogelius, Advances in bone preservation: identifying possible collagen preservation using sulfur speciation mapping, Palaeogeogr. Palaeoclimatol. Palaeoecol. 520 (2019) 181–187.
[43] V.L. Harvey, V.M. Egerton, A.T. Chamberlain, P.L. Manning, M. Buckley, Collagen fingerprinting: a new screening technique for radiocarbon dating ancient bone, PLoS One 11 (2016) 1–15, https://doi.org/10.1371/journal.pone.0150650.
[44] T.W. Stafford, K. Brendel, R.C. Duhamel, Radiocarbon, 13C and 15N analysis of fossil bone: removal of humates with XAD-2 resin, Geochim. Cosmochim. Acta 52 (1988) 2257–2267, https://doi.org/10.1016/0016-7037(88)90128-7.
[45] H. Bocherens, D.G. Drucker, D. Billiou, M. Patou-Mathis, B. Vandermeersch, Isotopic evidence for diet and subsistence pattern of the Saint-Césaire I Neanderthal: review and use of a multi-source mixing model, J. Hum. Evol. 49 (2005) 71–87.
[46] B.M. Lynch, R.W. Jefferies, A comparative analysis of the nitrogen content of bone as a means of establishing a relative temporal ordination of prehistoric burials, J. Archaeol. Sci. 9 (1982) 381–390.
[47] D.J. Ortner, M.S. Robinson, The effect of temperature on protein decay in bone: its significance in nitrogen dating of archaeological specimens, Am. Antiq. 37 (1972) 514–520.
[48] R.C. Dobberstein, M.J. Collins, O.E. Craig, G. Taylor, K.E.H. Penkman, S. Ritz-Timme, Archaeological collagen: why worry about collagen diagenesis? Archaeol. Anthropol. Sci. 1 (2009) 31–42.
[49] K. Douka, V. Slon, Z. Jacobs, C.B. Ramsey, M.V. Shunkov, A.P. Derevianko, F. Mafessoni, M.B. Kozlikin, B. Li, R. Grün, Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave, Nature 565 (2019) 640–644.
[50] Z. Jacobs, B. Li, M.V. Shunkov, M.B. Kozlikin, N.S. Bolikhovskaya, A. K. Agadjanian, V.A. Uliyanov, S.K. Vasiliev, K. O'Gorman, A.P. Derevianko, Timing of archaic hominin occupation of Denisova Cave in southern Siberia, Nature 565 (2019) 594–599.
[51] L.K. Horwitz, M. Chazan, Past and Present at Wonderwerk Cave, Northern Cape Province, South Africa, 2015.
[52] M. Chazan, F. Berna, J. Brink, M. Ecker, S. Holt, N. Porat, J.L. Thorp, L.K. Horwitz, Archeology, environment, and chronology of the Early Middle stone age component of wonderwerk cave, J. Paleolithic Archaeol. (2020), https://doi.org/10.1007/s41982-020-00051-8.
[53] K. Douka, S. Grimaldi, G. Boschian, A. del Lucchese, T.F.G. Higham, A new chronostratigraphic framework for the Upper Palaeolithic of Riparo Mochi (Italy), J. Hum. Evol. 62 (2012) 286–299.
[54] O. Wedage, A. Picin, J. Blinkhorn, K. Douka, S. Deraniyagala, N. Kourampas, N. Perera, I. Simpson, N. Boivin, M. Petraglia, P. Roberts, Microliths in the South Asian Rainforest ~45–4 Ka: New Insights from Fa-Hien Lena Cave, Sri Lanka, 2019, https://doi.org/10.1371/journal.pone.0222606.
[55] C. Fredericksen, M. Spriggs, W. Ambrose, Pamwak rockshelter: A Pleistocene site on Manus Island, Papua New Guinea, in: A. Smith, M. Spriggs, B. Fankhauser (Eds.), Sahul Rev. Pleistocene Archaeol. Aust. New Guinea Isl. Melanes, Prehistory Department, Australian National University Canberra, 1993, pp. 144–154.
[56] C. Gosden, N. Robertson, Models for Matenkupkum: interpreting a late pleistocene site from southern New Ireland, Papua New Guinea, in: J. Allen, C. Gosden (Eds.), Rep. Lapita Homel. Proj., Department of Prehistory, Research School of Pacific Studies, the Australian, 1991, pp. 20–45.
[57] J. Allen, C. Gosden, J.P. White, Human Pleistocene adaptations in the tropical island Pacific: recent evidence from New Ireland, a Greater Australian outlier, Antiquity 63 (1989) 548–561.
[58] J. Allen, The pre-Austronesian settlement of island Melanesia: implications for Lapita archaeology, Trans. Am. Philos. Soc. 86 (1996) 11–27.
[59] T.H.J. Niedermeyer, M. Strohalm, mMass as a software tool for the annotation of cyclic peptide tandem mass spectra, PLoS One 7 (2012), https://doi.org/10.1371/journal.pone.0044913.
[60] R Core Team, R: A Language and Environment for Statistical Computing, 2013.
[61] H. Wickham, ggplot2: Elegant Graphics For Data Analysis, Springer, 2016.
[62] M. Sielaff, J. Kuharev, T. Bohn, J. Hahlbrock, T. Bopp, S. Tenzer, U. Distler, Evaluation of FASP, SP3, and iST protocols for proteomic sample preparation in the low microgram range, J. Proteome Res. 16 (2017) 4060–4072, https://doi.org/10.1021/acs.jproteome.7b00433.
[63] S. Brown, S. Hebestreit, N. Wang, N. Boivin, K. Douka, K.K. Richter, S. Hebestreit, Zooarchaeology by Mass Spectrometry (ZooMS) for Bone Material - AmBiC Protocol, Protocols.Io, 2020, pp. 2–10, https://doi.org/10.17504/protocols.io.bffdjji6.
[64] S. Brown, S. Hebestreit, N. Wang, N. Boivin, K. Douka, K.K. Richter, S. Hebestreit, Zooarchaeology by Mass Spectrometry (ZooMS) for Bone Material - Acid Insoluble Protocol, Protocols.Io, 2020, pp. 1–12, https://doi.org/10.17504/protocols.io.bf43jqyn.
[65] S. Brown, S. Hebestreit, N. Wang, N. Boivin, K. Douka, K.K. Richter, S. Hebestreit, Zooarchaeology by Mass Spectrometry (ZooMS) for Bone Material - Acid Soluble Protocol, Protocols.Io, 2020, pp. 1–11, https://doi.org/10.17504/protocols.io.bf5bjq2n.
[66] N. Wang, S. Wilkin, S. Hebestreit, S. Brown, N. Boivin, K. Douka, K.K. Richter, S. Hebestreit, SP3 (Single-Pot, Solid-Phase, Sample-Preperation) Protocol for ZooMS applications, Protocols.Io Vol. 3, 2020, pp. 1–23, https://doi.org/10.17504/protocols.io.bf6pjrdn.
[67] C.S. Hughes, S. Moggridge, T. Müller, P.H. Sorensen, G.B. Morin, J. Krijgsveld, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments, Nat. Protoc. 14 (2019) 68–85, https://doi.org/10.1038/s41596-018-0082-x.
[68] S. Brown, K. Douka, M.J. Collins, K.K. Richter, On the standardization of ZooMS nomenclature, J. Proteome (5 November 2020), 104041.
[69] M. Buckley, S.W. Kansa, Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, south eastern Turkey, Archaeol. Anthropol. Sci. 3 (2011) 271–280, https://doi.org/10.1007/s12520-011-0066-z.
[70] J. Bradfield, T. Forssman, L. Spindler, A.R. Antonites, Identifying the animal species used to manufacture bone arrowheads in South Africa, Archaeol. Anthropol. Sci. 11 (2019) 2419–2434.
[71] M. Buckley, M. Gu, S. Shameer, S. Patel, A.T. Chamberlain, High-throughput collagen fingerprinting of intact microfaunal remains; a low-cost method for distinguishing between murine rodent bones, Rapid Commun. Mass Spectrom. 30 (2016) 805–812, https://doi.org/10.1002/rcm.7483.
[72] M. Buckley, M. Gu, J. Herman, J.A. Junno, C. Denys, A.T. Chamberlain, Species identification of voles and lemmings from Late Pleistocene deposits in Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting, Quat. Int. 483 (2018) 83–89, https://doi.org/10.1016/j.quaint.2018.03.015.
[73] B.O. Keller, J. Sui, A.B. Young, R.M. Whittal, Interferences and contaminants encountered in modern mass spectrometry, Anal. Chim. Acta 627 (2008) 71–81.
[74] M.W. Morley, P. Goldberg, V.A. Uliyanov, M.B. Kozlikin, M.V. Shunkov, A. P. Derevianko, Z. Jacobs, R.G. Roberts, Hominin and animal activities in the microstratigraphic record from Denisova Cave (Altai Mountains, Russia), Sci. Rep. 9 (2019) 1–12, https://doi.org/10.1038/s41598-019-49930-3.

[75] P. Goldberg, F. Berna, M. Chazan, Deposition and diagenesis in the earlier stone age of Wonderwerk cave, excavation 1, South Africa, Afr. Archaeol. Rev. 32 (2015) 613–643.
[76] I. Kontopoulos, K. Penkman, V.E. Mullin, L. Winkelbach, M. Unterländer, A. Scheu, S. Kreutzer, H.B. Hansen, A. Margaryan, M.D. Teasdale, B. Gehlen, M. Street, N. Lynnerup, I. Liritzis, A. Sampson, C. Papageorgopoulou, M.E. Allentoft, J. Burger, D.G. Bradley, M.J. Collins, Screening archaeological bone for palaeogenetic and palaeoproteomic studies, PLoS One 15 (2020) 1–17, https://doi.org/10.1371/journal.pone.0235146.
[77] T.P. Cleland, K. Voegele, M.H. Schweitzer, Empirical evaluation of bone extraction protocols, PLoS One 7 (2012) 1–9, https://doi.org/10.1371/journal.pone.0031443.

Supplementary Figures

Testing the efficacy and comparability of ZooMS protocols on archaeological bone

Wang, N., Brown, S., Ditchfield, P., Hebestreit, S., Shunkov, M., Kozilikin, M., Grimaldi, S., Chazan, M. Horwitz Kolska L., Spriggs, M., Summerhayes, G., Wedage, O., Richter Korzow, K., Douka, K.

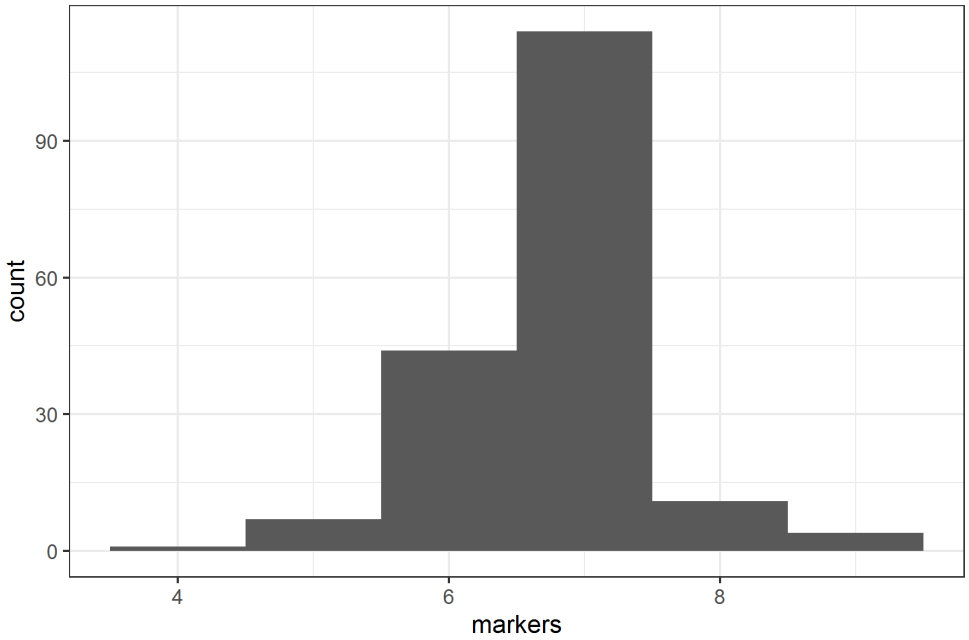


Figure S1. Distribution plot showing the number of ZooMS-specific diagnostic markers for 181 IDable samples following Protocol 1 (AmBic).

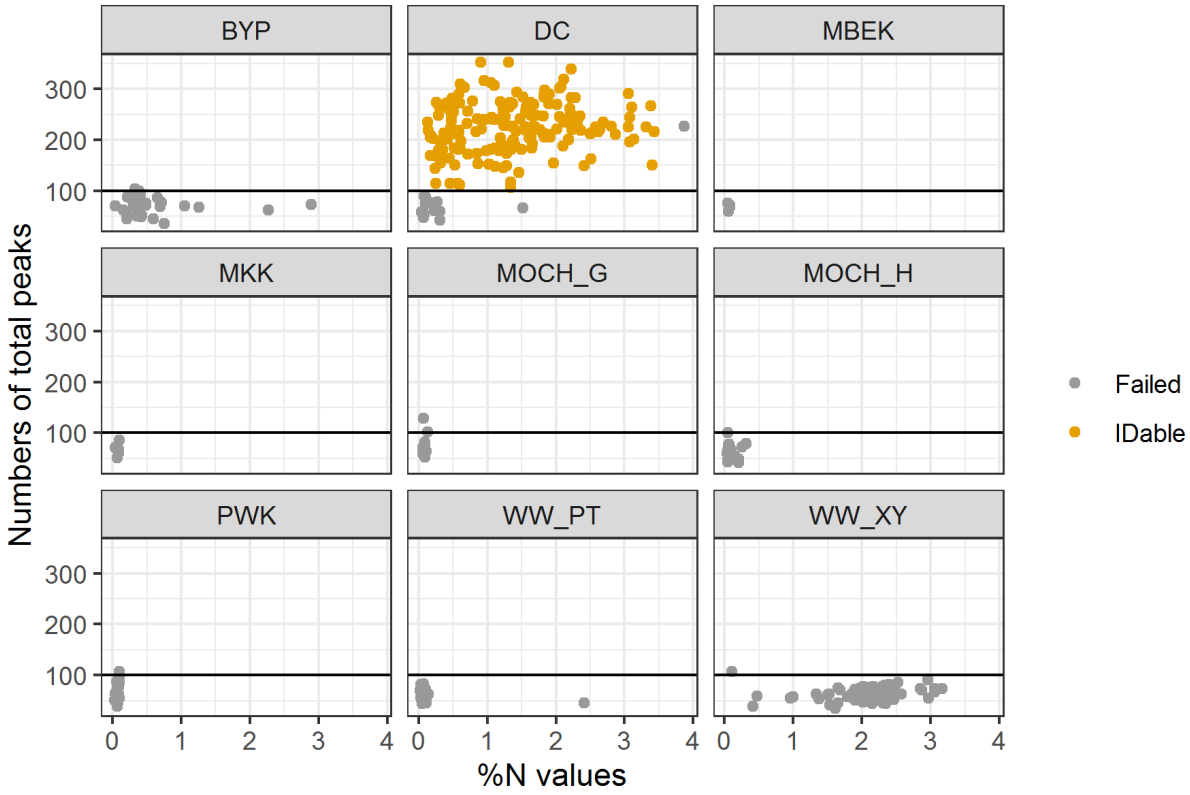


Figure S2. Correlation of number of total peaks and %N for the AmBic protocol per site. Each dot represents a unique sample and the horizontal line represents 100 peaks.

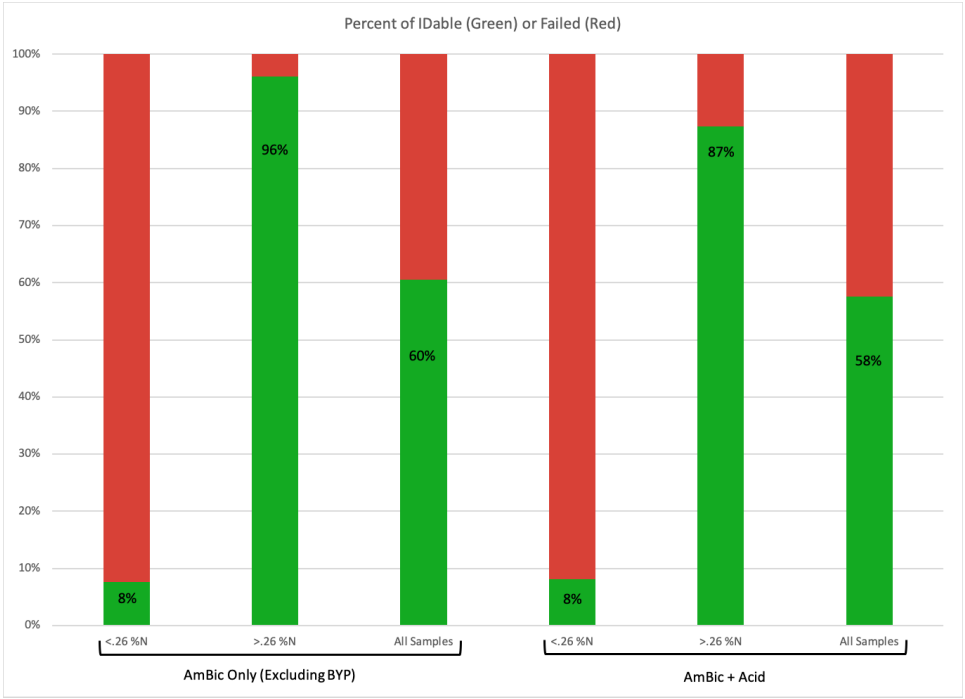


Figure S3. Percentage of IDable and failed samples when using 0.26% as a prediction threshold.

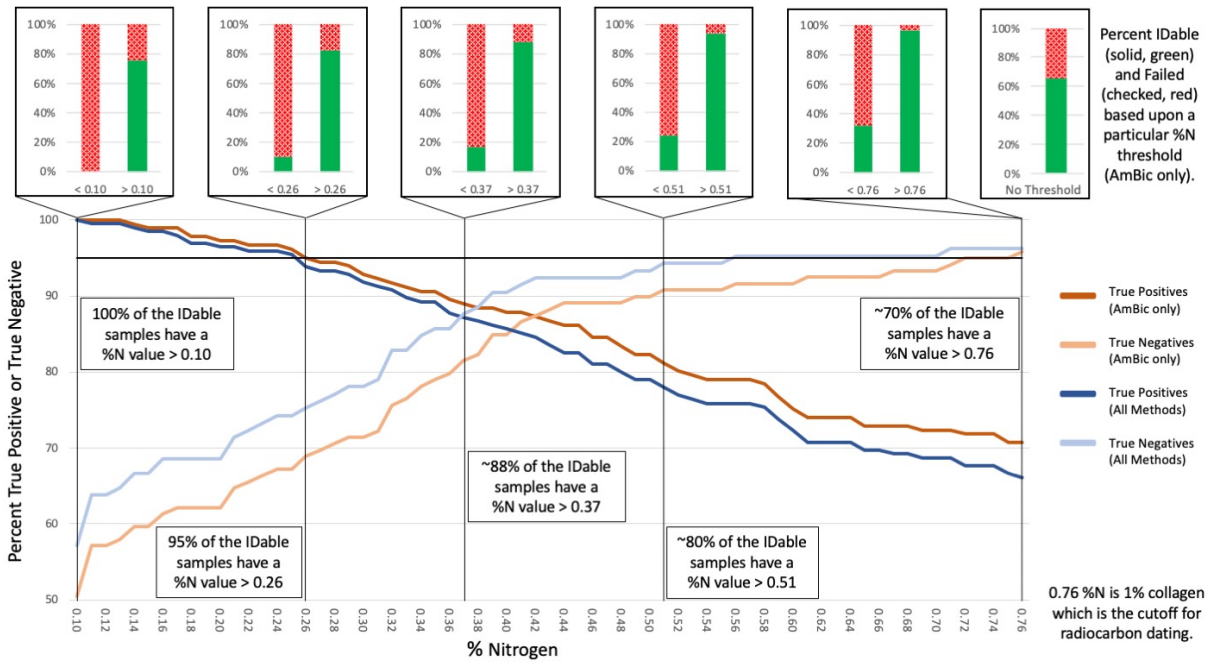


Figure S4. Comparison of thresholding at different %N values (excluding the samples from WW). The main graph shows the true positives (percent of IDable samples that have a %N value greater than the threshold, dark colors) and true negatives (percent of the fail samples that have a %N value lower than the threshold, light colors) for the AmBic method (brown) and all methods combined (blue). For key %N values, the graphs on top display the percentage of IDable and failed spectra that fall into each category given a specific threshold. The furthest right graph shows the percent IDable and failed spectra of the entire assemblage (n=300). 0.26 %N is chosen as a threshold with a true positive of 95% (solid line), meaning only 5% of the IDable spectra would be excluded by using this threshold. Lower or higher thresholds trade off percent success (green bars in the > %N value bar) for excluding IDable samples from analysis (true positives).

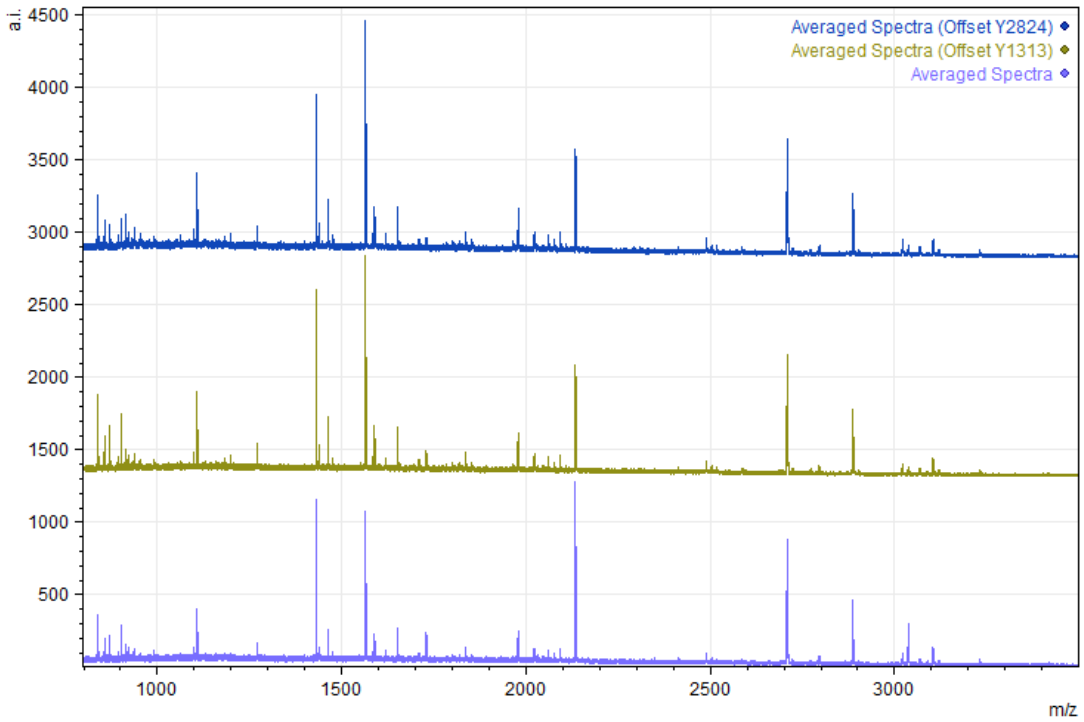


Figure S5. Spectra of the standard (positive control test) using the AmBic protocol for 3 weight groups (4-5 mg, 8-10 mg, 17-19 mg).

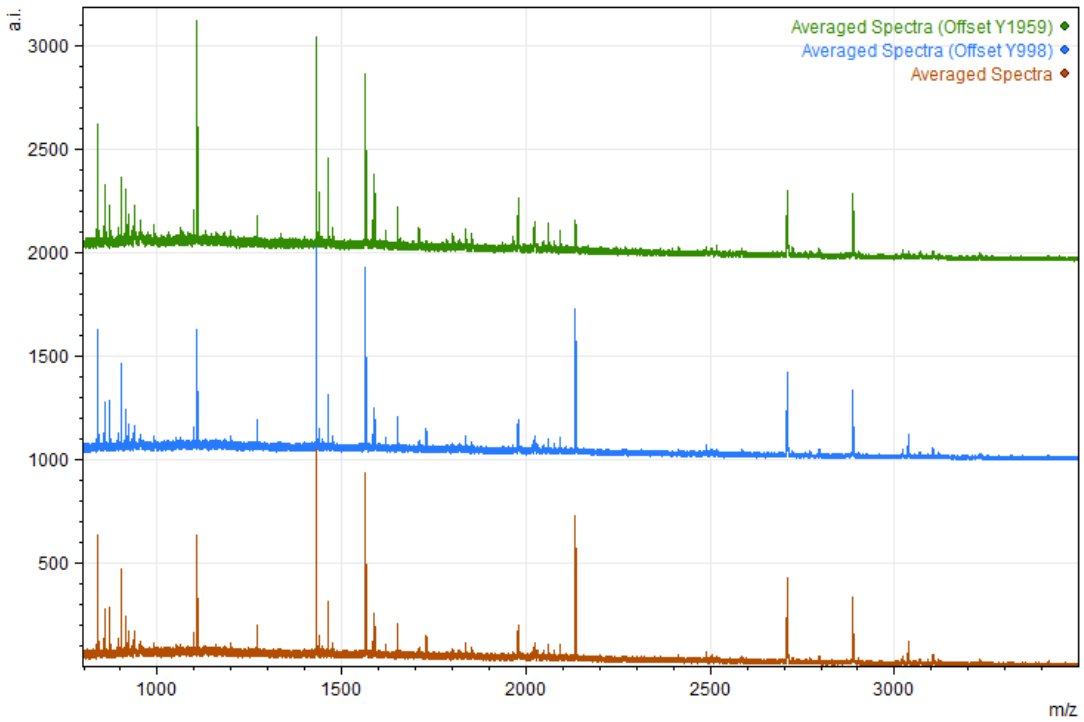


Figure S6. Spectra of the standard (positive control test) using the Acid-insoluble protocol for 3 weight groups (4-5 mg, 8-10 mg, 17-19 mg).

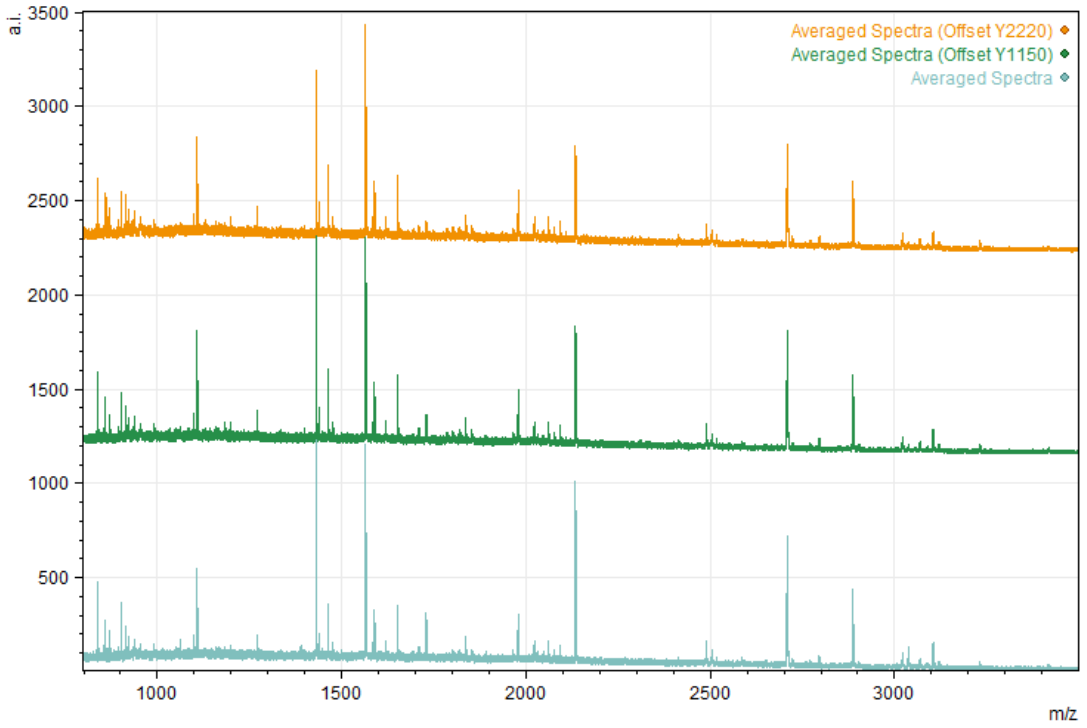


Figure S7. Spectra of the standard (positive control test) using the Acid-soluble protocol for 3 weight groups (4-5 mg, 8-10 mg, 17-19 mg).

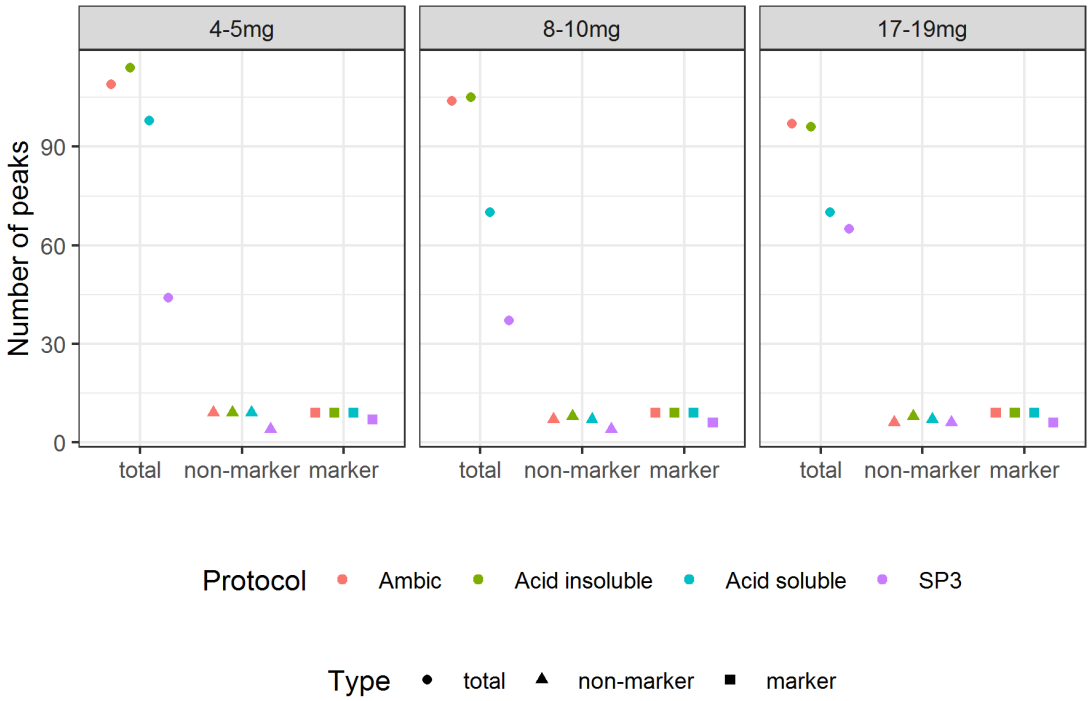


Figure S8. Peak counts for the standard (positive control test). The total peaks, non-marker and marker peaks were counted as described in the main text (2.1). One contaminant peak was found, this is not shown in the graph. The numerical format of this comparison is in Table S5.

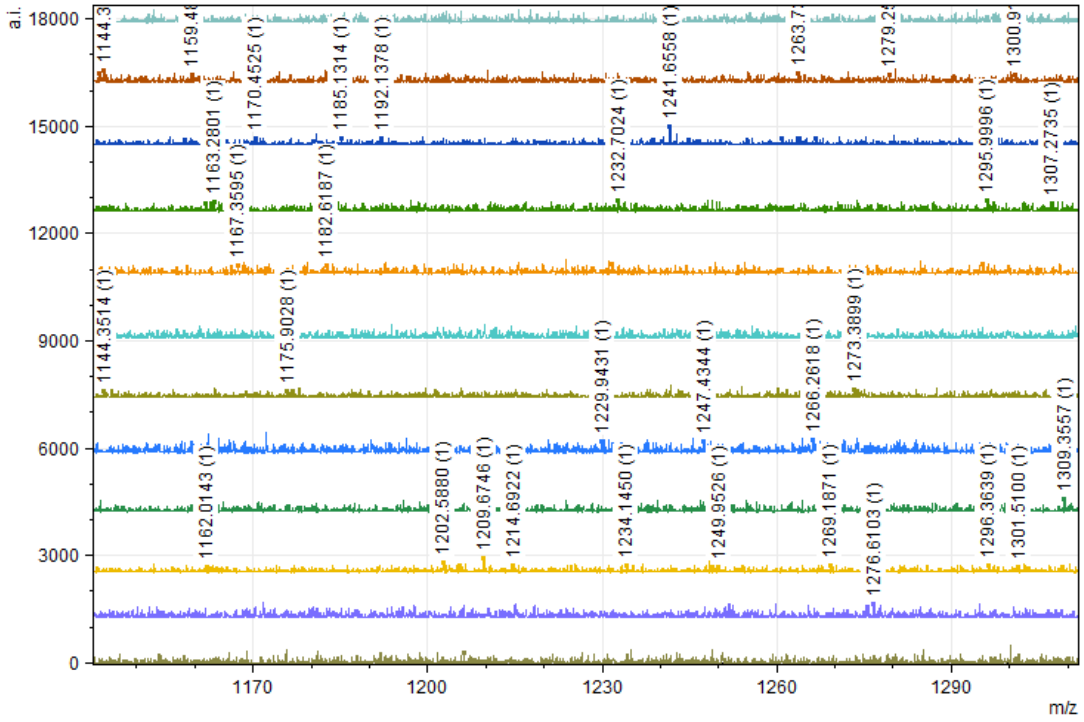


Figure S9. Spectra of 12 DC samples derived from the SP3 protocol, zoom-in to 1150-1300 m/z. Peak-picking s/n is 3.5. From top to down: DC5787, DC5785, DC5769, DC5770, DC5772, DC5775, DC5776, DC5778, DC5781, DC5782, DC5783, DC5788.

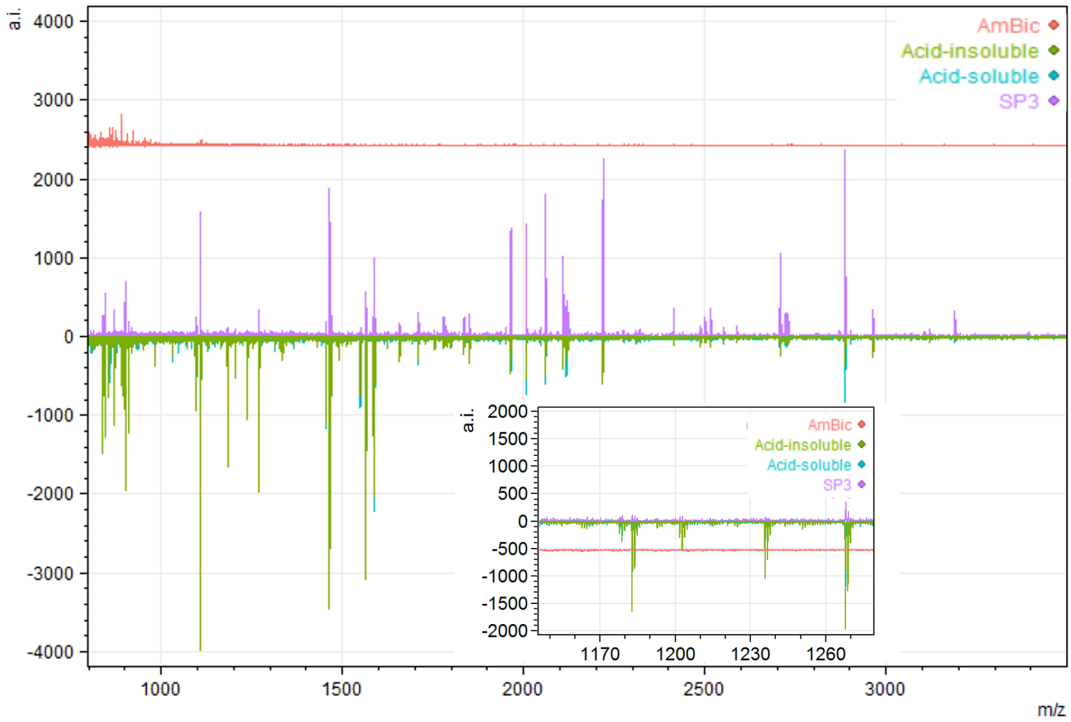


Figure S10. Comparison of the MALDI spectra for 4 protocols applied on the same sample (BYP20) (Ambic protocol in red, the Acid-based protocols in green and blue, SP3 in purple).

Manuscript B

Naihui Wang, Xu Yang, Zhuowei Tang, Cunding He, Xin Hu, Yinqiu Cui, Katerina Douka. "Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China." *Proceedings of the Royal Society B* 290, no. 2009 (2023): 20231129.

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Wang N, Xu Y, Tang Z, He C, Hu X, Cui Y, Douka K. 2023 Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China. *Proc. R. Soc. B* **290**: 20231129. https://doi.org/10.1098/rspb.2023.1129

Received: 23 May 2023
Accepted: 22 September 2023

Subject Category:
Palaeobiology

Subject Areas:
palaeontology

Keywords:
ZooMS, palaeoproteomics, deamidation, radiocarbon dating, Palaeolithic, camels

Authors for correspondence:
Naihui Wang
e-mail: nwang@gea.mpg.de
Yinqiu Cui
e-mail: cuiyq@jlu.edu.cn
Katerina Douka
e-mail: katerina.douka@univie.ac.at

†These authors contributed equally.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6875422.

THE ROYAL SOCIETY PUBLISHING

Large-scale application of palaeoproteomics (Zooarchaeology by Mass Spectrometry; ZooMS) in two Palaeolithic faunal assemblages from China

Naihui Wang[1,3,4,†], Yang Xu[1,†], Zhuowei Tang[2], Cunding He[5,6], Xin Hu[7], Yinqiu Cui[1] and Katerina Douka[3,8,9]

[1]School of Life Sciences, and [2]School of Archaeology, Jilin University, 130012 Changchun, People's Republic of China
[3]Max Planck Institute of Geoanthropology, 07745, Jena, Germany
[4]Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen, Germany
[5]China-Central Asia 'the Belt and Road' Joint Laboratory on Human and Environment Research, 710127 Xi'an, People's Republic of China
[6]School of Cultural Heritage, Northwest University, 710127 Xi'an, People's Republic of China
[7]Chongqing China Three Gorges Museum, 400013 Chongqing, People's Republic of China
[8]Department of Evolutionary Anthropology, Faculty of Life Sciences, and [9]Human Evolution and Archaeological Sciences (HEAS), University of Vienna, 1030 Vienna, Austria

NW, 0000-0003-2121-6030

The application of Zooarchaeology by Mass Spectrometry (ZooMS) on Pleistocene sites in Europe and northern Asia has resulted in the discovery of important new hominin fossils and has expanded the range of identified fauna. However, no systematic, large-scale application of ZooMS on Palaeolithic sites in East Asia has been attempted thus far. Here, we analyse 866 morphologically non-diagnostic bones from Jinsitai Cave in northeast China and Yumidong Cave in South China, from archaeological horizons dating to 150–10 ka BP. Bones from both sites revealed a high degree of collagen preservation and potentially time-related deamidation patterns, despite being located in very distinct environmental settings. At Jinsitai, we identified 31 camel bones, five of which were radiocarbon dated to 37–20 ka BP. All dated specimens correspond to colder periods of Marine Isotope Stages 3 and 2. We regard the presence of camels at Jinsitai as evidence of wild camels being a megafauna taxon targeted, most likely by early modern humans, during their expansion across northeast Asia. This large-scale application of ZooMS in China highlights the potential of the method for furthering our knowledge of the palaeoanthropological and zooarchaeological records of East Asia.

1. Introduction

Significant new archaeological and palaeoanthropological discoveries from East Asia have highlighted the region's importance in understanding late human evolution [1–5]. However, our knowledge of human presence and adaptation to these, often extreme, territories are limited, although some multi-period sites with long stratigraphies offer such potential (e.g. [6]).

Recent developments in ancient DNA (aDNA) research including the extraction of aDNA from sedimentary deposits and bone remains, have opened new exciting possibilities worldwide [7,8]. However, aDNA preservation is challenging in some regions of East Asia, particularly in warm and humid areas. Ancient proteins are an alternative group of biomolecules that often preserve

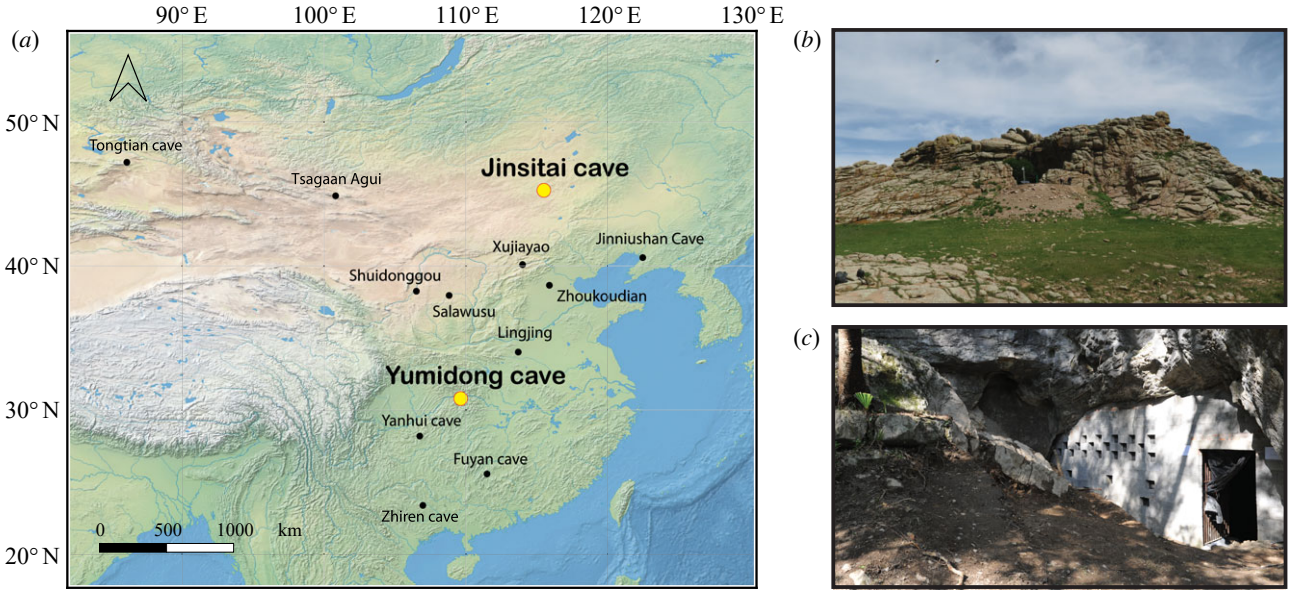


Figure 1. Map with location of sampled sites in this study. (*a*) Location of Yumidong Cave and Jinsitai Cave shown with yellow dots; (*b*) view towards Jinsitai Cave; (*c*) the entrance of Yumidong Cave. Base map from https://www.naturalearthdata.com/.

better and can help address research questions in palaeoanthropology and zooarchaeology [9,10]. Peptide mass fingerprinting of collagen, also known as zooarchaeology by mass spectrometry (ZooMS), is a powerful palaeoproteomic method for the taxonomic identification of collagenous materials such as bone, ivory and leather [11–14]. ZooMS involves the extraction of Type I collagen (COL1) and the generation of tryptic-digested peptide mass fingerprints using matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS). COL1, the major organic component (approx. 90%) in the bone of vertebrates, is a highly durable biomolecule, and peptides as old as 3.5 Myr have been extracted from bone remains [15].

ZooMS is particularly valuable analytical tool for screening highly fragmented bones that lack diagnostic features and therefore are not suitable for traditional zooarchaeological analyses [16–21]. The method performs well on bones from cold environments, while its success rates for bones from temperate, tropical and subtropical zones are generally lower [22].

In this work, we investigate the applicability of ZooMS in East Asia as part of a larger-scale, study involving numerous Pleistocene-age sites from across Eurasia (FINDER Project). The aims of our work in China were threefold. First, we wanted to examine whether the application of ZooMS on various Chinese sites—where the method had not been applied on a large scale before—would be successful and whether site location and age would be a major contributing factor to success or failure rates. Assuming a degree of successful collagen extraction, the second aim of this project revolved around the identification of new hominin remains and, finally, the third aim was an attempt to expand the morphology-based faunal identifications using ZooMS. While our initial goal was to include a large number of sites and bone material from different periods and depositional environments, the pandemic prevented us from studying a larger number of sites. Despite this limitation, this work, designed as a feasibility study for the recovery of ancient proteins from different locations in China, represents the largest application of ZooMS in East Asia to date.

2. Material and methods

We applied ZooMS to 866 unidentified bones from two Palaeolithic sites in China: Yumidong Cave in the south and Jinsitai Cave in the north (figure 1). Information about the sites, the analysed material and methods of analyses are detailed below.

(a) Studied sites

The two analysed sites are located in regions with distinct ecological and climatic settings, (assumed) biomolecular preservation conditions and research histories.

The Three Gorges region is a hub for archaeological and paleoanthropological research, with many sites having been discovered in recent decades. Yumidong Cave is a recently found karst cave in this region [23]. It consists of a large and nearly horizontal chamber, 70 m in length and 12–20 m in width. A 3 m in diameter vertical skylight provides air circulation and light, making the cave particularly attractive to human occupation. Excavations began in 2011 and focused on the area between the roof skylight and the cave entrance. Approximately 150 m² of surface area has been exposed thus far. The stratigraphy consists of a 6 m depth sequence divided in 18 distinct layers; the bedrock has not been reached (electronic supplementary material, appendix and figure S1). The excavations have yielded thousands of lithics and fauna remains, including 113 worked bones. Large limestone tools make up 97% of the lithic assemblage and belong to the cobble industry that prevailed in southern China during the Pleistocene. The faunal remains are attributed to the *Ailuropoda-Stegodon* fauna complex of Southeast Asia. Multiple dating methods have been applied to the site, and Bayesian analysis of 48 determinations established a geological and archaeological record spanning approximately 300 ka for Yumidong Cave [24].

Jinsitai Cave, located at the eastern end of the mid-latitude, semi-arid Eurasian belt, on the China-Mongolia border, is a rare Palaeolithic cave site with stratified sequence in northern China. The granite cave covers an area of nearly 120 m^2. Initial excavations in 2000–2001 depleted the deposit extensively, and subsequent excavations focused on the limited remaining sediment [24]. Around 5000 lithic artefacts, 3000 faunal remains and three hearths were discovered at the site, in nine stratigraphic layers (electronic supplementary material, appendix and figure S2). The upper layers contained a Late Upper

Palaeolithic assemblage of microblades and bifacially thinned points, alongside the traditional core-and-flake (small flakes) industry which is typical in contemporaneous sites in northern China. The lower layers were dominated by core-and-flake industry, while some distinctive Levallois flakes were described as Mousterian-like artefacts [25,26]. Some researchers regard the presence of this Mousterian-like industry at Jinsitai as evidence of a population dispersal or technological diffusion from the west. The lithic industry from the Mongolian site Tsagaan Agui was recently compared with the Jinsitai Levallois Mousterian [27] but more comparative techno-typological work needs to be done. The Jinsitai fauna is attributed to the *Mammuthus–Coelodonta* faunal complex, although no mammoths are included in the assemblage. Radiocarbon dating on bone collagen suggests human occupation of the cave from around 47–44 ka BP until the Holocene [26].

(b) Materials

For Yumidong Cave, we randomly selected 121 non-diagnostic bone fragments (no teeth or antler) from layer 2 to layer 9, all of which were excavated in 2013. We limited our sampling to bones from the uppermost Middle Pleistocene and Late Pleistocene layers, due to concerns of collagen preservation. The average size of the sampled bone fragments was approximately 4 cm.

For Jinsitai Cave, we analysed all 745 unidentifiable bones from the site. They were excavated during the 2000–2001 fieldwork but lack exact contextual information. This is because while all bones were collected and grouped during excavation by layers, after the zooarchaeological analysis, fragments lacking morphological characteristics from every layer were mixed together. We used this mixed 'unidentified' assemblage for our ZooMS work. During sampling, we noted the presence of glue on the bones from Jinsitai, verified as polyvinyl acetate applied to the bones shortly after excavation. The glue has aged, cracked and concealed possible modifications on the bone surface. The analysed specimens, most of which were long shafts, varied in size, and we recorded their weight before sampling.

(c) Sampling and data generating

Each bone was subsampled using a circular diamond saw blade. To eliminate surface contaminants such as glue and sediment, a small area of the bone was sandblasted before removing a chip of approximately 20 mg for ZooMS analysis, or approximately 600 mg for radiocarbon dating.

We used the ZooMS acid-insoluble protocol [22,28] for 866 samples. Seven bones from Jinsitai were submitted to the Oxford Radiocarbon Accelerator Unit and were dated using routine ultrafiltration methodologies [29]. More details can be found in the electronic supplementary material, appendix.

(d) Data processing

The calculation of glutamine deamidation is based on Wilson *et al*. [30]. The amino acid glutamine (Q) in collagen peptides may undergo *post-mortem* deamidation, resulting in a mass shift of 0.984 Da. COL1α1 508–519 (GVQGPPGPAGPR) (marker P1 or cet1 from previous research [31]) contains a single glutamine site identified at *m*/*z* 1105.5 (non-deamidated) and *m*/*z* 1106.5 (deamidated). Theoretically, deamidation values range from 0 to 1. A value of 1 indicates no or negligible deamidation in COL1α1 508–519 peptides, while 0 indicates nearly complete deamidation. Values greater than 1 may also be observed due to baseline noise, which can distort the relative intensity.

MALDI-TOF spectra were converted from t2d files to mzXML files using T2D converter [32] and processed using the mMass 5.5.0 [33]. Previously published COL1 peptide markers were used for ZooMS-based taxonomic identifications [15,18,28,34–36]. The raw radiocarbon data were calibrated to calendar years using OxCal v. 4.4.4 [37] and the IntCal20 calibration curve [38]. Statistical analysis and visualization were conducted in R [39] with the ggplot2 package [40].

3. Results and discussion

(a) Deamidation

To investigate the influence of local environmental conditions and age on bone collagen preservation and overall ZooMS performance, we analysed the proteomic profiles using glutamine deamidation observed in the peptide COL1α1 508–519. This peptide sequence is conserved across mammalian species and has been used in previous studies as a proxy for the relative 'thermal age' of samples or to detect intrusive bones of different ages in an *in situ* deposit [41,42].

Deamidation values for both sites deviated from a normal distribution (electronic supplementary material, appendix and figure S3–S5). Therefore, we used non-parametric Kernel density estimation to assess the overall deamidation patterns (figure 2, insert). The two sites had distinct deamidation patterns. The average deamidation value at Yumidong was lower and less variable than Jinsitai. Despite their wide variation, the median deamidation value for the Jinsitai dataset was 0.62, whereas the value for Yumidong was significantly lower at approximately 0.15. This suggests that, in general, bones from Jinsitai were less deamidated, which agrees well with the higher ZooMS identification rate (see next section), as well as the younger overall age of Jinsitai. Based on the published chronology for each site, the deepest layer at Jinsitai post-dates 50 ka BP. Therefore, the entire Jinsitai deposit corresponds only to the upper part of layer 2 (63–14 ka BP) of Yumidong (electronic supplementary material, figures S1 and S2) [24,26]. To explore further whether the deamidation patterns in both sites were linked to time, we plotted the deamidation values against the ZooMS taxa of Jinsitai (labels in yellow) and the stratigraphic layers of Yumidong (labels in purple) on the same figure (figure 2). While most taxa in Jinsitai presented broad and overlapping ranges of deamidation, an indirect time-related pattern can be observed when comparing the deamidation values of *Sus* sp. (pig/wild boar) and Rhinocerotidae (woolly rhinoceros), whose deamidation ranges hardly overlap. The two taxa are thought to be separated temporally at the site. Zooarchaeological study of the Jinsitai fauna reported pigs/wild boars ($n = 2$) exclusively in layer 2, whereas woolly rhinoceros ($n = 123$) were only found in layers 3–8 (data in electronic supplementary material, appendix, table S2) [43]. Layer 2 corresponds to the Holocene, as evidenced by pottery sherds found there, while woolly rhinoceros pre-date the Holocene and are believed to have gone extinct in East Asia around the Allerød oscillation approximately 13 ka [25,44]. The outlier JST 285 (triplicate) in the Rhinocerotidae group in figure 2 suggests a well-preserved specimen, possibly of a younger age. In Yumidong, most of the deamidation variation occurred in the two upper layers, while deamidation levels in layers 4–9 were close to 0. There was one exception, YMD 113 (Cervidae/Antilopinae) from layer 9, which exhibited a deamidation value of

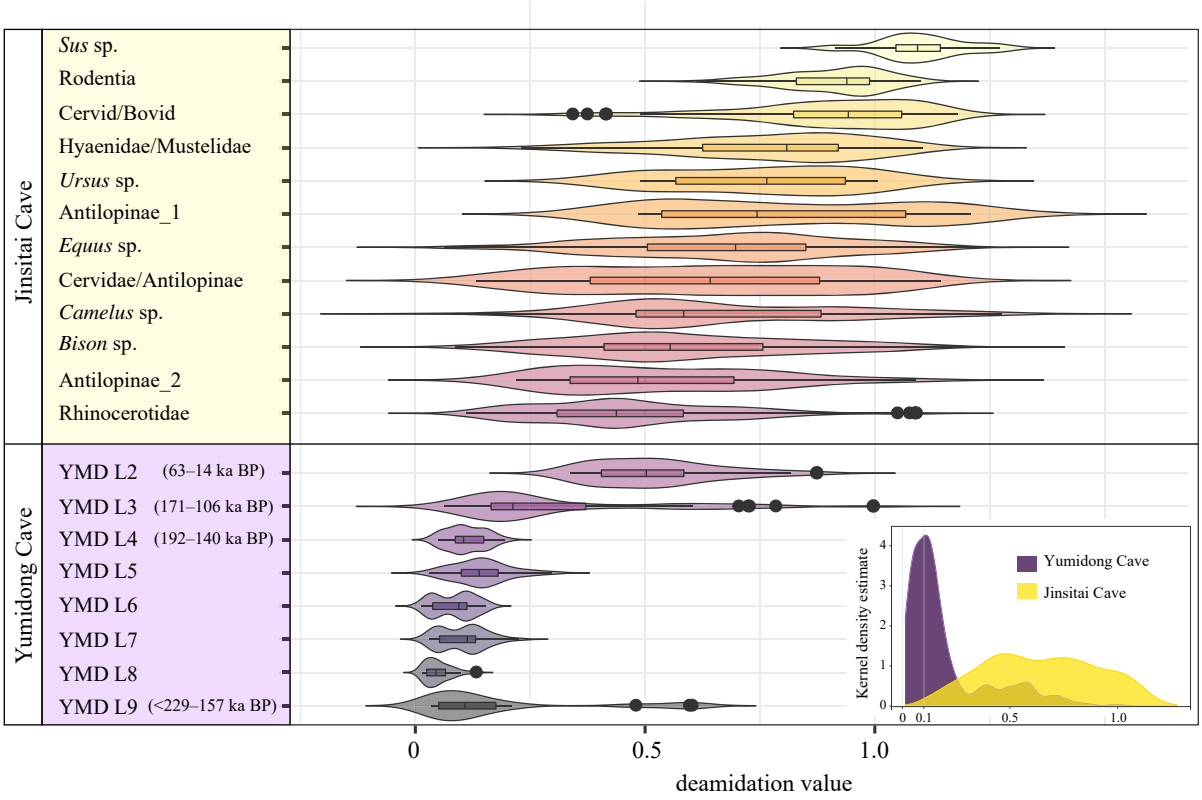


Figure 2. Visualization of deamidation levels at Yumidong and Jinsitai. Insert panel: kernel density estimate of deamidation value on ZooMS identifiable mammals. Main figure: violin plot on deamidation levels for Jinsitai (in yellow, around 47–44 ka BP to the Holocene) grouped by ZooMS-identified taxa, and Yumidong (in purple, Middle and Late Pleistocene) grouped by archaeological layers. The plotted data for Jinsitai includes *Sus* sp., $n = 4$; Rodentia, $n = 28$; Cervid/Bovid, $n = 47$; Hyaenidae/Mustelidae, $n = 40$; *Ursus* sp., $n = 4$; Antilopinae_1, $n = 9$; *Equus* sp., $n = 189$; Cervidae/Antilopinae, $n = 53$; *Camelus* sp., $n = 31$; *Bison* sp., $n = 125$; Antilopinae_2, $n = 19$; Rhinocerotidae, $n = 121$, shown in descending order on the basis of their median deamidation values. The Yumidong dataset includes ZooMS identifiable specimens from layer 2, $n = 13$; layer 3, $n = 16$; layer 4, $n = 9$; layer 5, $n = 10$; layer 6, $n = 15$; layer 7, $n = 15$; layer 8, $n = 14$; layer 9, $n = 8$, totally 100 (data in appendix, electronic supplementary material, table S1). Chronological data for Yumidong from [24].

around 0.5. This outlier might indicate either an intrusion or extraordinarily well-preserved collagen.

The wide range of deamidation values for the glutamine of COL1α1 508–519 in bones from Jinsitai and in layers 2 and 3 of Yumidong suggests that deamidation is still ongoing in these deposits. By contrast, bones from layers 4 to 9 in Yumidong are fully deamidated. This agrees with the data that these layers are considerably older, probably pre-dating 140 ka.

While deamidation values cannot be used directly as an indicator of age, in some cases they can serve as a relative age indicator for chronologically separated fauna groups within a single site. However, it is important to note that the deamidation process is influenced by both diagenetic and laboratory-induced factors [45]. Therefore, we ought to stress that our findings are specific to the sampled deposits. The results of this study support previous research suggesting that the deamidation may be considered an indicator of collagen preservation and a thermal age proxy among different fossil assemblages. However, achieving chronological resolution in absolute terms is extremely challenging—if not impossible [41,42,46,47].

(b) ZooMS taxonomic results and comparison with zooarchaeological data

Despite the antiquity and location of Yumidong and Jinsitai caves in the subtropical and the temperate zones of East Asia, respectively, the ZooMS-based identification rates were unexpectedly high. Out of the 745 bones analysed from Jinsitai Cave, 90% had enough collagen for assignment to the order or genus level. The success rate for Yumidong Cave is 83%. To assess if the new data fit within the overall zooarchaeological record for each site, we compared the ZooMS-based identifications with the morphological identifications.

In Yumidong Cave, 21 of the 121 analysed samples failed to yield enough collagen. No significant correlation was observed between stratigraphic depth and success of ZooMS identification (electronic supplementary material, appendix and figure S6), possibly due to the limited number of samples included in this study.

A total of 1530 bones were recovered from the first 15 layers of Yumidong, with the majority coming from layers 2, 5, 10 and 11. About one-third of the faunal assemblage (480 specimens) was identified morphologically, revealing a high diversity of taxa (approx. 40) which included *Stegodon* sp., *Cervus* sp., *Muntiacus muntjac* (southern red muntjac), Caprinae, *Bubalus* sp. or *Bos* sp., *Sus scrofa* (wild boar), *Equus* sp. (horse), *Stephanorhinus* sp. (two-horned rhinoceros), *Megatapirus augustus* (giant tapir), *Ursus thibetanus* (Asian black bear), *Ailuropoda melanoleuca* (giant panda), a few carnivore taxa and microfauna [48]. Based on the abundance of cervids (31%) and stegodons (23%), the Yumidong fauna belongs to the *Stegodon–Ailuropoda* faunal complex, a typical fauna complex of large-bodied mammals widely distributed from East Asia to mainland Southeast Asia during the Late Pleistocene [49–51].

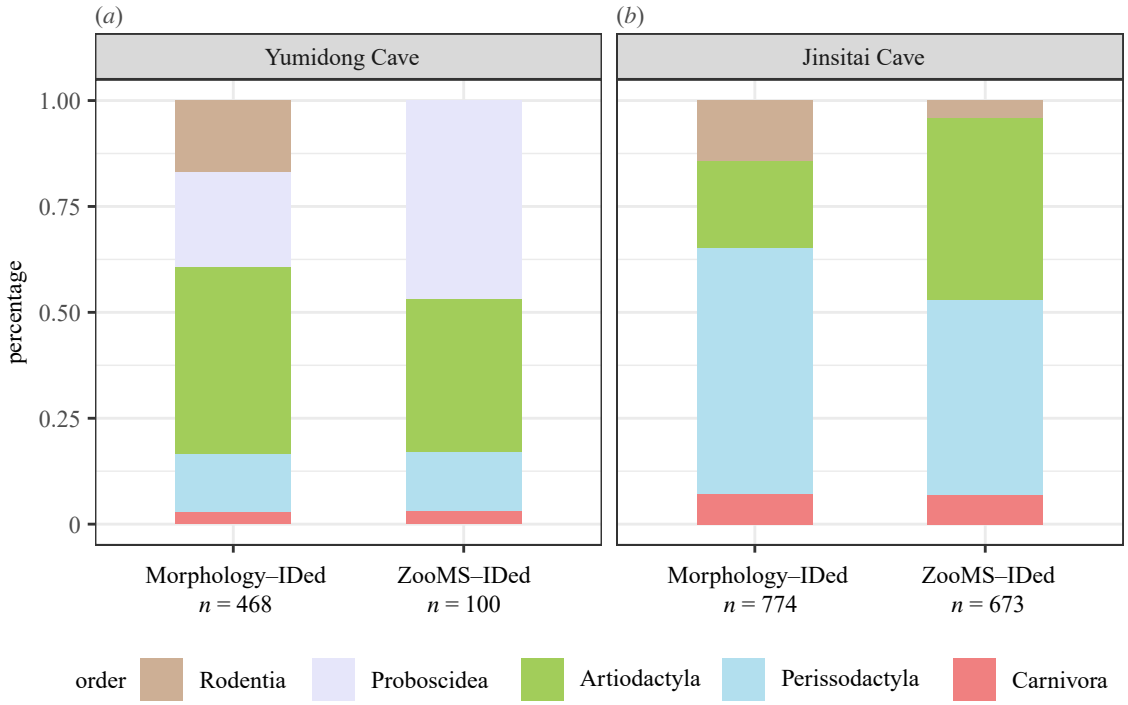


Figure 3. ZooMS results compared with zooarchaeological data for (*a*) Yumidong and (*b*) Jinsitai caves. For each site, the bar plot indicates the percentage of identified specimens of mammals based on morphology and on ZooMS. Further details are provided in the electronic supplementary material, appendix and tables S2–S4.

Prior to comparing morphological versus ZooMS datasets, we removed 12 bones of microfauna taxa from the morphological dataset, resulting in a total of 468 morphologically identified mammals [48] (electronic supplementary material, appendix and table S2). We included rodents in our comparison because they account for about 17% of the morphological dataset from Yumidong, contributing to half of the species diversity at the site.

The Yumidong ZooMS results reveal a reduced diversity of mammals compared with the morphological data (electronic supplementary material, appendix and table S2). This may be explained by the fact that (i) the ZooMS dataset ($n = 100$) is nearly five times smaller than the morphologically identified one ($n = 468$); (ii) the ZooMS-analysed bones have an average size of 4 cm, thus most microfauna would have been excluded during sampling; and (iii) the low resolution in separating cervids and bovids using ZooMS could mask the overall taxonomic diversity.

In order to compare the two datasets, we classified the morphologically identified mammals and the ZooMS-identified mammals into five orders (Rodentia, Proboscidea, Artiodactyla, Perissodactyla and Carnivora) (figure 3*a*).

The abundance of carnivores (e.g. hyenas) in caves is often used to determine whether hominins or carnivores were the driving force for the accumulation of an assemblage [52–54]. The percentage of Carnivora, around 3%, is similar in both datasets of Yumidong, falling below the 20% threshold required for designating a fauna assemblage as carnivore accumulation. This confirms published work on bone and lithic artefacts analysis that highlights the dominant role of hominins in the formation of the site [55,56].

The Perissodactyla group includes extinct regional species, such as two-horned rhinoceros, giant tapirs and very few horses. Rhinos and tapirs share all diagnostic peptide markers [57], leading to a combined Ceratomorpha category in our ZooMS dataset. Interestingly, the percentage of order Perissodactyla is consistent in both ZooMS and morphological datasets (14%) (figure 3*a*).

The Proboscidea group is the most abundant in our ZooMS results. Stegodon, the typical species in the *Stegodon–Ailuropoda* faunal complex in Southern China, represents 22% of the morphological assemblage at Yumidong Cave. Stegodon remains, primarily consisting of cranial and foot elements [58] have been found in all layers at Yumidong, and over 85% were neonate and juvenile individuals. While our current ZooMS reference library lacks stegodon, we identified 47 bones whose spectra matched the Elephantidae ZooMS fingerprint [35]. Since *Elephas* coexisted with extinct stegodons in southern China throughout the Pleistocene, the ZooMS-identified proboscideans from Yumidong are assigned to Elephantoidae, a group that includes both Stegodontidae and Elephantidae. Elephantoidae represents 47% of the ZooMS assemblage, making it the most abundant taxon in the Yumiding ZooMS dataset.

The zooarchaeological studies [48,56] suggest diverse strategies for the exploitation of large animals. The inhabitants of Yumidong scavenged or hunted stegodons but only transferred the skulls and limbs of neonate and juvenile individuals back to the cave. Two stegodon tusks from layers 2 and 5 were modified for the production of tools. The remains of two-horned rhinoceros show a bias towards older individuals and less preference on transporting body elements to the site.

Artiodactyla (mainly cervids and bovids) is the largest group in the morphological dataset but ranks second after Elephantoidae in the ZooMS results. Bovids, *Bos* or *Bubalus*, account for 13% in the morphological dataset but 31% in the ZooMS data. ZooMS can separate *Bos* from *Bubalus*, which is challenging morphologically. However, ZooMS cannot reach genus-level identification for cervids (including *Cervus* sp. and southern red muntjac at the site). The fragmented ZooMS assemblage shows a larger percentage of *Bos* and *Bubalus*, while the

proportion of cervid specimens decreases significantly (29% versus 4%). This disparity may be due to body size, with large animals better represented in the ZooMS assemblage. Two cervid antlers were modified into tools [56].

Despite the limited and exploratory nature of the application of ZooMS at Yumidong, our analysis provides a new perspective on the highly fragmented bones from the site. It complements the morphological identifications that dominantly rely on teeth. The least abundant groups (Carnivora and Perissodactyla) in the morphological dataset (3% and 14%) align with the ZooMS-based dataset, while the more dominant orders (Rodentia, Artiodactyla and Perissodactyla) show differences in the ZooMS dataset that may be the result of body-size effects.

In Jinsitai, out of 745 samples analysed using ZooMS, 68 had no collagen. The remaining 677 samples were successfully analysed, with one identified as Aves (bird) and 673 as mammals (electronic supplementary material, appendix and table S4). Three bones were assigned to an 'unknown' category due to unmatched peptide masses.

While all analysed bones lacked diagnostic morphological features, not all of them were small in size. A slight correlation (Cohen's $d = 0.31$) was found between bone weights and ZooMS success rate, with rates of 93% for the greater than 10 g group, 91% for the 3–10 g group and 88% for the less than 3 g group, resulting in an overall identification rate of 90% (electronic supplementary material, appendix and figure S7).

The main excavation of Jinsitai Cave yielded 2372 bones from layers 2 to 8. In total, 778 (33%) specimens were identified morphologically to genus or species level. More than half of the morphologically identified specimens (51%) come from layer 4, while layers 2, 3 and 5 each yielded 11–13%. In addition to four bird bones, 15 mammalian taxa were identified at the site, including *Myospalax aspalax* (zokor), *Marmota bobak* (bobak marmot), *Cervus elaphus* (red deer), *Procapra przewalskii*, *Pachygazella* sp., *Spirocerus* sp., *Bison* sp. (bison), *Equus ferus przewalskii* (Przewalski horse), *Equus hemionus* (Asiatic wild ass), *Sus scrofa* (pig/wild boar), *Coelodonta antiquitatis* (woolly rhinoceros), *Ursus spelaeus* (cave bear), *Crocuta crocuta ultima* (hyena), *Canis lupus* (wolf) and *Gulo* sp. (wolverine). Minimum numbers of individuals (MNIs) were estimated for these taxa (electronic supplementary material, appendix and table S2) [43].

The Jinsitai fauna is attributed to the *Mammuthus–Coelodonta* faunal complex despite the absence of mammoth. The site is located in a relatively open landscape compared with eastern regions where mammoths have been recorded [59,60]. Although no comparable cave site exists in the region, similar taxa, with the exception of cave bear and bobak marmot, were found at the open-air site of Salawusu in Inner Mongolia [61] (figure 1). Sediment pollen analysis suggested a shift from a taiga-steppe to a less-cold steppe ecosystem during the human occupation of Jinsitai Cave.

To compare morphological and ZooMS-identified taxa, the 774 morphologically identified mammals [43] were categorized into four orders (Rodentia, Artiodactyla, Perissodactyla and Carnivora), and they were compared with the 673 ZooMS-identified mammals, also grouped into four groups (figure 3*b*).

The Carnivora category in the ZooMS and the morphological datasets both represent approximately 7% of the assemblage at Jinsitai, mainly made by cave bears and hyenas. Among the morphologically identified specimens, around 160 bones showed traces of burning, over 140 had cut-marks, and less than 40 had signs of carnivore gnawing [43]. Since only one mustelid was identified morphologically, we hypothesize that the indistinguishable taxon Hyaenidae/Mustelidae in ZooMS mostly contains hyenas.

In the morphological dataset, Rodentia accounted for about 14% of the Jinsitai assemblage, represented by bobak marmots ($n = 108$) and zokors ($n = 3$). However, in the ZooMS assemblage, only 4% ($n = 28$) of the bones were assigned to rodents, and they were almost exclusively found in the smallest weight group (less than 3 g). Although bobak marmot and zokor were not present in the ZooMS reference library, the Jinsitai rodents showed closest match to the alpine marmot (CDS: XP_015350976.1). The lower number of rodents in the ZooMS dataset suggests that rodents were not severely fragmented. The deamidation level of rodents indicates a relatively late appearance at Jinsitai, consistent with previous zooarchaeological studies on bobak marmots, which were limited to layers 2 to 4 at Jinsitai and may have been the result of burrowing activity [62]. Bobak marmot is absent at Salawusu, in the same region [43,61].

Artiodactyla is the most diverse group in the Jinsitai faunal assemblage and includes pigs/wild boars, red deer, four bovids (*Procapra przewalskii*, *Pachygazella* sp., *Spirocerus* sp., *Bison*. sp.) [43] and the newly identified *Camelus* sp. (camel) (see text below). Two pig/wild boar remains were morphologically identified both in layer 2, and their presence was confirmed by ZooMS, albeit very infrequent ($n = 4$). Bison accounted for 9% ($n = 71$) of the morphological dataset, but this value doubled in the ZooMS data (18%, $n = 121$), indicating a potentially higher fragmentation level for this taxon. Although morphological and ZooMS analyses could not determine the Jinsitai bison remains to species level, it has been suggested that all Late Pleistocene bison remains in the northeast China plain should be identified as *Bison priscus* (steppe bison) due to the lack of reliably identified alternatives [63]. Red deer is the only morphologically identified cervid at Jinsitai ($n = 25$), with various axial and appendicular elements, as well as four antler fragments. By contrast, the remaining three local Antilopinae taxa, *Procapra przewalskii*, *Pachygazella* sp. (extinct) and *Spirocerus* sp. (extinct), were exclusively identified by horn fragments ($n =$ 17, 43 and 1, respectively), which are their most distinctive parts. Almost half of the identified horns had cut-marks at the roots, resulting from the removal activity on crania [43]. Using ZooMS, we were not able to identify any cervid or Antilopinae bones to the genus level at Jinsitai, due to their phylogenetic closeness and absence in the ZooMS reference library. Instead, 129 specimens were grouped into five 'ZooMS taxa' based on distinct marker combinations (for more details see the electronic supplementary material, appendix). Of these, the 'Cervid/Bovid' group ($n = 47$) represented the most generic assignment due to the lack of one or two ZooMS markers. The remaining four groups ($n = 82$) each represented a combination of seven markers, suggesting the presence of at least four ZooMS unidentified species at Jinsitai. The ambivalent classification of cervids hinders further discussion on the exploitation of cervids or antelopes at Jinsitai.

The order Perissodactyla includes two *Equus* species (Przewalski horse and Asiatic wild ass) and woolly rhinoceros at Jinsitai. Woolly rhinoceros were equally represented in the morphological and ZooMS datasets (16% versus 18%). Based on morphological identification [43], the two

Equus species were the most abundant taxa, accounting for 42% of the Jinsitai fauna, with 220 Przewalski horse specimens representing 12 individuals and 106 Asiatic wild ass specimens representing 13 individuals. Over 90% of identified elements were teeth and distal extremities, suggesting a possible preference by hominins for transporting skull and lower limbs to the site. Evidence from other sites in adjacent regions [64] shows that Przewalski horse/Asiatic wild ass was a substantial food resource for hominins. Using the ZooMS trypsin-digestion protocol, the Przewalski horse and Asiatic wild ass were indistinguishable and were identified as a combined *Equus* taxon. *Equus* sp. accounted for 28% of the ZooMS assemblage, much less than the morphological dataset. The over-representation of *Equus* in the morphological identifications may be attributed to the distinct morphology of horse teeth [65,66]. At Jinsitai, 67% of the morphologically identified *Equus* remains were teeth (n = 219). Interestingly, none of these teeth was from calves or young adults (less than 4 years) [43]; instead, the horse composition suggests a long-term exploitation of prime and old adults at Jinsitai.

The comparison between Jinsitai morphological and ZooMS datasets reveals specific differences. For example, cervids and bovids are more abundant in the ZooMS data, while rodents and *Equus* are less frequently found. The two assemblages represent the entire fauna collection from the main excavations of the site. Although the ZooMS data from Jinsitai presented here lack stratigraphic context, the deamidation analysis detected a few taxa (pig/wild boar and rodents) of which the presence at the site was relatively short and recent. Furthermore, four different ZooMS marker combinations were identified on cervids/bovids, representing at least four species. To achieve a more detailed taxonomic resolution, expanding the ZooMS reference with bovid species found in East Asia is necessary to clarify the new marker combinations [36].

(c) Camels in Jinsitai Cave

An unexpected discovery was the identification of camel remains in the faunal assemblage of Jinsitai. Thirty-one (n = 31) camel bone fragments were discovered using ZooMS (spectra in electronic supplementary material, appendix and figure S8), comprising 5% of the ZooMS-identified dataset. While ZooMS can identify two extant species in the genus *Camelus* at the species level (*C. bactrianus* and *C. dromedarius*) [15], the extinct *C. knoblochi* is not included in the current reference library. It is likely that the *C. knoblochi* shares most, if not all, ZooMS markers with the *C. ferus* (wild bactrian camel). The extinct 'giant' camel *C. knoblochi* was part of the *Mammuthus–Coelodonta* faunal complex that inhabited Asia for tens of thousands of years, although the exact date of its extinction remains uncertain [67].

Traditionally, camels are not considered a targeted species for Eurasian Palaeolithic hunter groups, and their remains are rarely found at cave sites. Camel skeletal remains preserve more diagnostic features than other megafauna species, hence their morphological identification should be, in principle, easy to achieve. Their absence, therefore, may be either because their feral predecessors were not numerous in the landscape thus rarely targeted, or, when hunted, transportation of body parts between killing sites and camping sites was limited. Following ZooMS identification, the 31 camel bones were morphologically examined; none preserved diagnostic features and all were heavily fractured. Seventeen fragments were from long bones, three from flat bones and four from irregular bones (electronic supplementary material, appendix and table S4). Like most taxa in the Jinsitai ZooMS assemblage, the size of camel fragments varied considerably, and were equally identified in the less than 3 g and greater than 10 g groups. Two camel long bone fragments showed possible traces of burning, probably due to heating to a low temperature since bone collagen was still present. The highly fragmentary nature of the camel bones discovered at Jinsitai may suggest that the camel bones underwent extensive level of modification and damage pre- and post-deposition. It is possible that while humans exploited camel remains, they only transferred specific body parts to the cave. This could explain the absence of more diagnostic bones (teeth, crania and vertebrae).

To establish the absolute timing of camel presence at Jinsitai, we radiocarbon dated seven ZooMS-identified camel bones, all of which were dense and/or large fragments. Of these, five produced enough collagen for dating. The results indicate that the five dated specimens represent at least four individuals (electronic supplementary material, appendix, table S5 and figure S9). OxA-X-3115-12 (JST 244) was produced on a low collagen bone and we cannot rule out that this is a minimum age. Notwithstanding, camel bones were deposited at Jinsitai during distinct periods, at 20.5 ka cal BP, at 26 ka cal BP, at 31 ka cal BP and at approximately 36 ka cal BP (ka cal BP = calendar thousand years before present). When plotted against the Greenland oxygen isotope record (NGRIP) [68] these new dates fall in cold conditions of the marine isotope stages (MIS) 3 and 2, particularly Greenland stadials 2.1, 3, 5 and 8 [69] (figure 4). The chronometric data we report here clearly indicates that camel presence at Jinsitai was not ephemeral or a one-off encounter. Instead, it suggests targeted and repeated exploitation of this animal for at least 17 millennia.

The deamidation levels of the dated camel bones were also examined. Four of the five specimens showed deamidation values ranging from 0.4 to 0.6. The exception was JST 628 (approx. 26 ka cal BP), with an average deamidation value of 0.95, indicating a nearly non-deamidated profile. This specimen was not the oldest among the directly dated bones. The inconsistency between radiocarbon age and deamidation level cautions against using deamidation as a molecular 'clock'.

In recent years, camel remains have been identified and reported at various Palaeolithic localities in Eurasia and Africa (figure 5). In Western Asia and North Africa, most instances appear to belong to the *Camelus thomasi* (wild dromedary) based on findings from mostly open-air sites in Egypt [71], Sudan [72] and Syria [73], dating to the Middle and Late Pleistocene. In Siberia, camel aDNA has been extracted at Denisova Cave from Middle Palaeolithic sediments dating to 140–120 ka BP [74].

Camelid bones have been reported from early Upper Palaeolithic sites in Uzbekistan (Samarkandskaya) [75], Siberia (Kamenka 1) [76] and Mongolia (Otson Tsokhio) [77,78]. These probably belong to the two-humped wild bactrian camel although there is still uncertainty as to its relationship and time of extinction with *Camelus knoblochi* [67]. Remains of *C. knoblochi*, the largest Eurasian two-humped member of the genus *Camelus*, have been recently

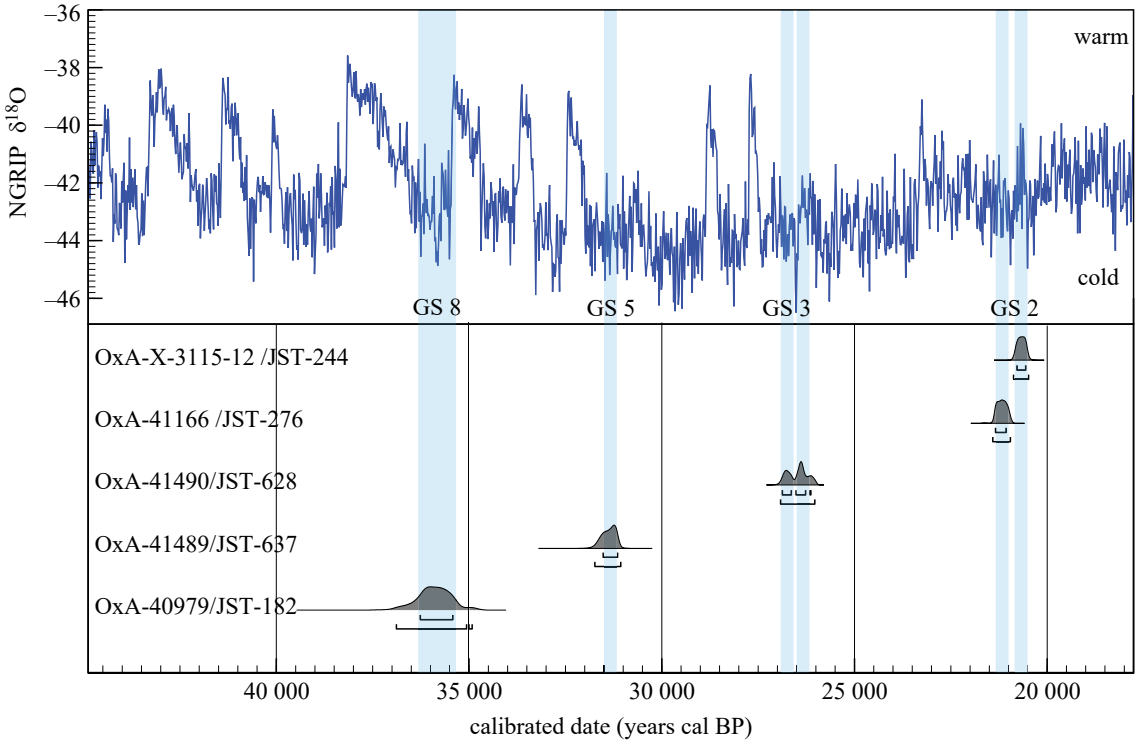


Figure 4. Calibrated probability functions of the radiocarbon ages for five camel bones from Jinsitai (data in electronic supplementary material, appendix, table S5). The raw ages were calibrated using OxCal v. 4.4.4 [37] and the IntCal20 terrestrial curve [38]. The Greenland oxygen isotope record (NGRIP) is shown on top of the graph. The blue shaded lines represent the 68.2% probability ranges of the calibrated radiocarbon ages and appear to correspond well to the cold Greenland stadials [69].

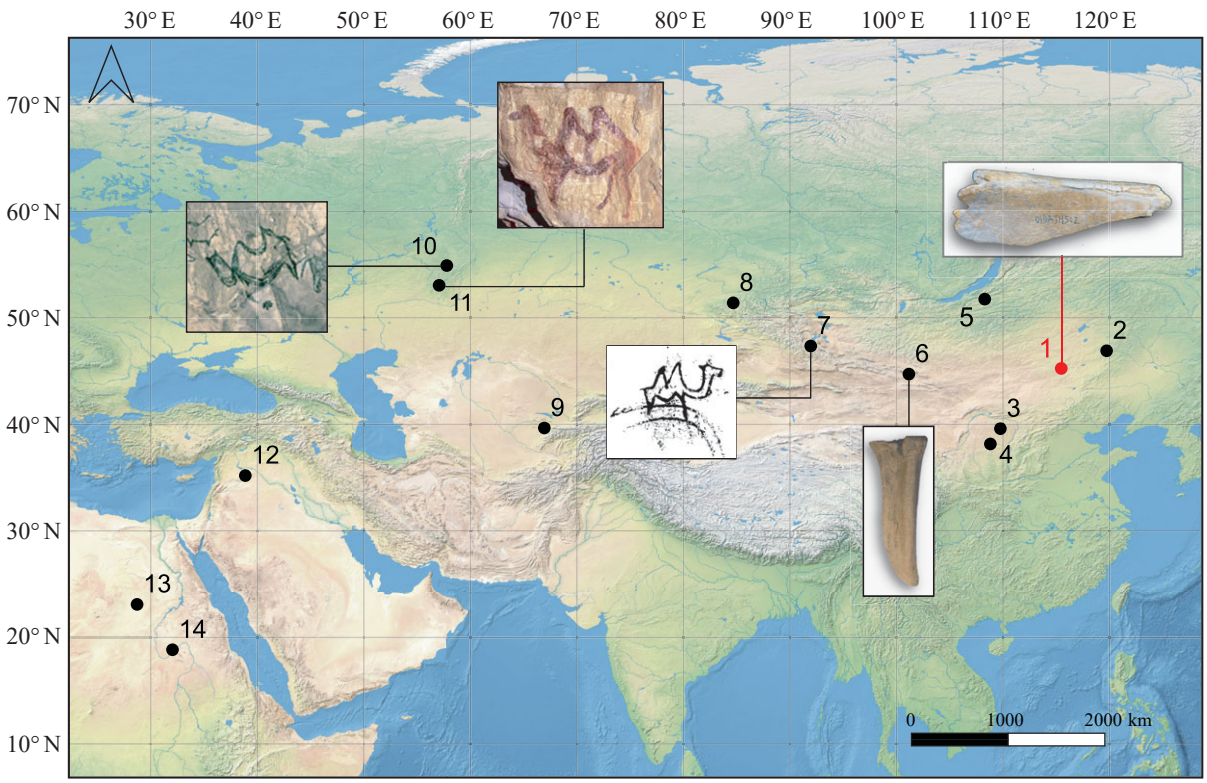


Figure 5. Camel bone remains and camel rock art depictions from Palaeolithic localities. 1, Jinsitai (China); 2, Otson Tsokhio (Mongolia); 3, Wulanmulun (China); 4, Salawusu (China); 5, Kamenka A (Russia); 6, Tsagaan Agui Cave (Mongolia); 7, Khoid Tsenkheriin Agui Cave (Mongolia); 8, Denisova Cave (Russia); 9, Samarkandskaya (Uzbekistan); 10, Ignatievskaya Cave (Russia); 11, Kapova Cave (Russia); 12, Nadaouiyeh Aïn Askar (Syria); 13, Bir Tarfawi (Egypt); 14, HP766 Wadi Umm Rahau (Sudan). Images of rock art provided by V. Shirokov; redrawing of camel rock art from Khoid Tsenkheriin Agui Cave and camel metacarpal from Tsagaan Agui, both modified after [70]. Base map from https://www.naturalearthdata.com/.

reported from Tsagaan Agui Cave in southern Mongolia [70]. In China, *C. knoblochi* bones have been previously identified at the palaeontological locality Dabusu, possibly dating to around 20 ka [79], and a few undated specimens were reported from two Palaeolithic sites, Wulanmulun [80] and Salawusu [81] in Inner Mongolia.

Interestingly, rare instances of Northern Asian parietal cave art found in the southern Urals in Russia (Kapova

Cave or Shulgan-Tash, and Ignatievskaya Cave) and Mongolia (Khoid Tsenkheriin Agui Cave), depict two-humped camels among other taxa [82–84] (figure 5). The Urals sites are estimated to date to approximately 20–15 ka BP, while Khoid Tsenkher Cave is believed to pre-date the Last Glacial Maximum. An interesting scene of camel hunting engraved on a mammoth tusk found near River Tom in West Siberia has been dated to minimum 13 ka BP [85].

Among these sporadic occurrences of camel bones and camel depictions, Jinsitai has the most numerous and well-dated camel remains so far. Camel presence there spanned at least 17 millennia. While it is not possible to specify which hominin species targeted camels at each site, it seems that both archaic (Asian Neanderthals and Denisovans perhaps, e.g. at Denisova Cave), as well as early modern humans, were exploiting this taxon. As hunter–gatherer populations expanded across North and Central Asia, they encountered camelids among the diverse megafauna. The extinction of the giant wild camel, *Camelus knoblochi*, probably occurred around 20 ka BP [79], which aligns well with the age determinations at Jinsitai for JST 244 and JST 276. However, further research at other sites is necessary to establish last appearance dates for the species with any confidence. Ongoing studies on the genetic profile of the camel bones discovered at Jinsitai aim to provide a better understanding of the evolutionary history of wild camels in Asia.

4. Conclusion

ZooMS has emerged as a valuable biomolecular tool which complements and enhances the traditional zooarchaeological research, especially when dealing with highly fragmented faunal assemblages. In this study, we conducted the first systematic large-scale application of ZooMS in China, analysing nearly 900 bones from two Palaeolithic sites. The analysis of glutamine deamidation at Jinsitai revealed an ongoing deamidation process with a possible temporal correlation to specific taxa. By contrast, at the much older layers of Yumidong, we observed nearly complete deamidation in bones dating back to 140–106 ka BP or before. We successfully extracted bone collagen from Yumidong layers dating back as far as 150 ka, in the subtropical zone of southern China. This opens up exciting possibilities for large-scale screening for collagen and other biomolecules in Pleistocene bones from such latitudes.

We identified 31 camel bones at Jinistai Cave, a previously unknown taxon at the site. Five of these bones were radiocarbon dated to between 37 and 20 ka cal BP; their punctuated presence at the site so far falls in stadial conditions of MIS 3 and MIS 2. The presence of camels at Jinsitai during colder periods provides significant insights into the broad spectrum of animal exploitation performed by early groups of, most likely, modern humans as they spread across the vast ranges of northeast Asia. Such findings highlight the advantages of using novel analytical tools, such as ZooMS, to study non-diagnostic bone assemblages.

Our work highlights the need for the broader application of ZooMS and other biomolecular approaches in East Asia. As ZooMS is applied to new regions, further developmental work is necessary. The current ZooMS reference library contains mostly, if not exclusively, North Eurasian taxa which limits meaningful identification and comparison of results. Given the high success rates, we report here, enlarging the reference library with extinct and extant East Asian taxa and the broader application of collagen fingerprinting to more archaeological assemblages promise exciting results for the future.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All ZooMS spectra files have been deposited in Mendeley data. ('ZooMS data of Yumidong and Jinsitai caves', Mendeley Data, V1, https://doi.org/10.17632/ssf27rywhh.2)

Supplementary material is available online [86].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. N.W.: data curation, visualization, writing—original draft, writing—review and editing; Y.X.: data curation, validation, writing—original draft, writing—review and editing; Z.T.: investigation, resources; C.H.: investigation, resources; X.H.: resources; Y.C.: conceptualization, methodology, supervision; K.D.: conceptualization, data curation, project administration, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. The authors declare no competing interests.

Funding. This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program to K.D. (grant agreement 715069-FINDER-ERC-2016). N.W. received doctoral funding through the FINDER project. Y.X. received research funding from the Science and Technology Development Project of Jilin Province (grant no. 20210508040RQ).

Acknowledgements. We thank Feng Li (Peking University) for providing a photograph of Jinsitai Cave, Vladimir N. Shirokov (Russian Academy of Sciences) for providing camel images from Ignatievskaya and Kapova Caves, Arina Khatsenovich (Russian Academy of Sciences) for the camel bone image from Tsagaan Agui and parietal art from Khoid Tsenkheriin Agui Caves, and Noel Amano (Max Planck Institute of Geoanthropology) for valuable discussions during the course of this work.

References

1. Bae CJ, Douka K, Petraglia MD. 2017 On the origin of modern humans: Asian perspectives. *Science* **358**, aai9067. (doi:10.1126/science.aai9067)
2. Shackelford L *et al.* 2018 Additional evidence for early modern human morphological diversity in Southeast Asia at Tam Pa Ling, Laos. *Quat. Int.* **466**, 93–106. (doi:10.1016/j.quaint.2016.12.002)
3. Groucutt HS *et al.* 2018 *Homo sapiens* in Arabia by 85 000 years ago. *Nat. Ecol. Evol.* **2**, 800–809. (doi:10.1038/s41559-018-0518-2)
4. Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. 2021 Origins of modern human ancestry. *Nature* **590**, 229–237. (doi:10.1038/s41586-021-03244-5)
5. Liu W, Wu X. 2022 Morphological diversities and evolutionary implications of the Late Middle Pleistocene hominins in China. *Acta Anthropol. Sin.* **41**, 563.
6. Zhang Y, Stiner MC, Dennel R, Wang C, Zhang S. 2010 Zooarchaeological perspectives on the Chinese Early and Late Paleolithic from the Ma'anshan site (Guizhou, south China). *J. Archaeol. Sci.* **37**, 2066–2077. (doi:10.1016/j.jas.2010.03.012)
7. Epp LS, Zimmermann HH, Stoof-Leichsenring KR. 2019 Sampling and extraction of ancient DNA from sediments. *Methods Mol. Biol.* **1963**, 31–44. (doi:10.1007/978-1-4939-9176-1_5)
8. Orlando L *et al.* 2021 Ancient DNA analysis. *Nat. Rev. Methods Primers* **1**, 1–26. (doi:10.1038/s43586-020-00011-0)

9. Chen F *et al.* 2019 A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature* **569**, 409–412. (doi:10.1038/s41586-019-1139-x)
10. Welker F. 2018 Palaeoproteomics for human evolution studies. *Quat. Sci. Rev.* **190**, 137–147. (doi:10.1016/j.quascirev.2018.04.033)
11. 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 Commun. Mass Spectrom.* **23**, 3843–3854. (doi:10.1002/rcm.4316)
12. Fiddyment S *et al.* 2015 Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting. *Proc. Natl Acad. Sci. USA* **112**, 15 066–15 071. (doi:10.1073/pnas.1512264112)
13. Coutu AN, Whitelaw G, Le Roux P, Sealy J. 2016 Earliest evidence for the ivory trade in Southern Africa: isotopic and ZooMS analysis of seventh–tenth century AD ivory from KwaZulu-Natal. *African Archaeol. Rev.* **33**, 411–435. (doi:10.1007/s10437-016-9232-0)
14. Richter KK, Codlin MC, Seabrook M, Warinner C. 2022 A primer for ZooMS applications in archaeology. *Proc. Natl Acad. Sci. USA* **119**, e2109323119. (doi:10.1073/pnas.2109323119)
15. Rybczynski N, Gosse JC, Harington CR, Wogelius RA, Hidy AJ, Buckley M. 2013 Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Commun.* **4**, 1550. (doi:10.1038/ncomms2516)
16. 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. *J. Archaeol. Sci.* **54**, 279–286. (doi:10.1016/j.jas.2014.12.010)
17. Brown S *et al.* 2016 Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6**, 23559. (doi:10.1038/srep23559)
18. Welker F *et al.* 2016 Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl Acad. Sci. USA* **113**, 11 162–11 167. (doi:10.1073/pnas.1605834113)
19. Devièse T *et al.* 2017 Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition. *Proc. Natl Acad. Sci. USA* **114**, 10 606–10 611. (doi:10.1073/pnas.1709235114)
20. 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). *Sci. Rep.* **9**, 12350. (doi:10.1038/s41598-019-48706-z)
21. Martisius NL *et al.* 2020 Non-destructive ZooMS identification reveals strategic bone tool raw material selection by Neandertals. *Sci. Rep.* **10**, 7746. (doi:10.1038/s41598-020-64358-w)
22. Naihui W *et al.* 2021 Testing the efficacy and comparability of ZooMS protocols on archaeological bone. *J. Proteomics* **233**, 104078. (doi:10.1016/j.jprot.2020.104078)
23. Wei G *et al.*. 2020 Recent discovery of a unique Paleolithic industry from the Yumidong Cave site in the Three Gorges region of Yangtze River, southwest China. *Quat. Int.* **434**, 107–120.
24. Shao Q, Philippe A, He C, Jin M, Huang M, Jiao Y, Voinchet P, Lin M, Bahain J-J. 2022 Applying a Bayesian approach for refining the chronostratigraphy of the Yumidong site in the Three Gorges region, central China. *Quat. Geochronol.* **70**, 101304. (doi:10.1016/j.quageo.2022.101304)
25. Wang X, Wei J, Chen Q, Tang Z, Wang C. 2010 A preliminary study on the excavation of the Jinsitai cave site. *Acta Anthropol. Sin.* **29**, 15–30.
26. Li F *et al.* 2018 The easternmost Middle Paleolithic (Mousterian) from Jinsitai Cave, North China. *J. Hum. Evol.* **114**, 76–84. (doi:10.1016/j.jhevol.2017.10.004)
27. Khatsenovich AM, Rybin EP, Tserendagva Y, Bazargur D, Margad-Erdene G, Marchenko DV, Gunchinsuren B, Olsen JW, Derevianko AP. 2023 The Middle Paleolithic of Tsagaan Agui Cave in the Gobi Altai region of Mongolia and its Siberian and Central Asian links. *Archaeol. Res. Asia* **35**, 100462. (doi:10.1016/j.ara.2023.100462)
28. 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 Commun. Mass Spectrom.* **23**, 3843–3854. (doi:10.1002/rcm.4316)
29. Brock F, Higham T, Ditchfield P, Ramsey CB. 2010 Current pretreatment methods for AMS radiocarbon dating at the Oxford radiocarbon accelerator unit (ORAU). *Radiocarbon* **52**, 103–112. (doi:10.1017/S0033822200045069)
30. Wilson J, Van Doorn NL, Collins MJ. 2012 Assessing the extent of bone degradation using glutamine deamidation in collagen. *Anal. Chem.* **84**, 9041–9048. (doi:10.1021/ac301333t)
31. Buckley M, Fraser S, Herman J, Melton ND, Mulville J, Pálsdóttir AH. 2014 Species identification of archaeological marine mammals using collagen fingerprinting. *J. Archaeol. Sci.* **41**, 631–641. (doi:10.1016/j.jas.2013.08.021)
32. Gao Y. 2016 *T2D converter | raw2ms*. See https://pepchem.org/download/converter.html.
33. Strohalm M, Kavan D, Novák P, Volný M, Havlícek V. 2010 mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Anal. Chem.* **82**, 4648–4651. (doi:10.1021/ac100818g)
34. Buckley M, Kansa SW. 2011 Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. *Archaeol. Anthropol. Sci.* **3**, 271–280. (doi:10.1007/s12520-011-0066-z)
35. Buckley M, Larkin N, Collins M. 2011 Mammoth and mastodon collagen sequences; survival and utility. *Geochim. Cosmochim. Acta* **75**, 2007–2016. (doi:10.1016/j.gca.2011.01.022)
36. Janzen A *et al.* 2021 Distinguishing African bovids using zooarchaeology by mass spectrometry (ZooMS): new peptide markers and insights into Iron Age economies in Zambia. *PLoS ONE* **16**, e0251061. (doi:10.1371/journal.pone.0251061)
37. Ramsey B. 2021 *OxCal v.4.4.4*. See http://c14.arch.ox.ac.uk/oxcal.
38. Reimer PJ *et al.* 2020 The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757. (doi:10.1017/RDC.2020.41)
39. R Core Team. 2013 *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.
40. Wickham H. 2016 *ggplot2: elegant graphics for data analysis*. Cham, Switzerland: Springer International Publishing.
41. Van Doorn NL, Wilson J, Hollund H, Soressi M, Collins MJ. 2012 Site-specific deamidation of glutamine: a new marker of bone collagen deterioration. *Rapid Commun. Mass Spectrom.* **26**, 2319–2327. (doi:10.1002/rcm.6351)
42. Chowdhury MP, Wogelius R, Manning PL, Metz L, Slimak L, Buckley M. 2019 Collagen deamidation in archaeological bone as an assessment for relative decay rates. *Archaeometry* **61**, 1382–1398. (doi:10.1111/arcm.12492)
43. Luo P. 2007 *The research on the fauna and ancient ecological environment of the Late Pleistocene in Jinsitai cave site*. Jilin University. See https://cdmd.cnki.com.cn/Article/CDMD-10183-2007105490.htm.
44. Keliang Z, Haitao J, Yuan W, Haowen T, Yaping Z, Junyi GE, Xinying Z, Changzhu JIN, Xiaoqiang LI. 2022 New radiocarbon evidence on the woolly mammoth and rhinoceros in China. *Acta Anthropol. Sin.* **41**, 551.
45. Simpson JP, Penkman KEH, Demarchi B, Koon H, Collins MJ, Thomas-Oates J, Shapiro B, Stark M, Wilson J. 2016 The effects of demineralisation and sampling point variability on the measurement of glutamine deamidation in type I collagen extracted from bone. *J. Archaeol. Sci.* **69**, 29–38. (doi:10.1016/j.jas.2016.02.002)
46. Schroeter ER, Cleland TP. 2016 Glutamine deamidation: an indicator of antiquity, or preservational quality? *Rapid Commun. Mass Spectrom.* **30**, 251–255. (doi:10.1002/rcm.7445)
47. Brown S, Kozlikin M, Shunkov M, Derevianko A, Higham T, Douka K, Richter KK. 2021 Examining collagen preservation through glutamine deamidation at Denisova Cave. *J. Archaeol. Sci.* **133**, 105454. (doi:10.1016/j.jas.2021.105454)
48. He C. 2016 *The lithic industry and human behaviour at Yumidong site*. Jilin University. See https://cdmd.cnki.com.cn/Article/CDMD-10183-1017014181.htm.
49. Colbert EH, Hooijer DA. 1953 Pleistocene mammals from the limestone fissures of Szechwan, China. *Bull. Am. Mus. Nat. Hist.* **102**, 1–134.
50. Louys J, Curnoe D, Tong H. 2007 Characteristics of Pleistocene megafauna extinctions in Southeast Asia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **243**, 152–173. (doi:10.1016/j.palaeo.2006.07.011)

51. Fan Y, Shao Q, Bacon A-M, Liao W, Wang W. 2022 Late Pleistocene large-bodied mammalian fauna from Mocun cave in south China: palaeontological, chronological and biogeographical implications. *Quat. Sci. Rev.* **294**, 107741. (doi:10.1016/j.quascirev.2022.107741)
52. Cruz-Uribe K. 1991 Distinguishing hyena from hominid bone accumulations. *J. Field Archaeol.* **18**, 467–486.
53. Enloe JG, David F, Baryshnikov G. 2000 Hyenas and hunters: zooarchaeological investigations at Prolom II Cave, Crimea. *Int. J. Osteoarchaeol.* **10**, 310–324. (doi:10.1002/1099-1212(200009/10)10:5<310::AID-OA562>3.0.CO;2-B)
54. Pickering TR. 2002 Reconsideration of criteria for differentiating faunal assemblages accumulated by hyenas and hominids. *Int. J. Osteoarchaeol.* **12**, 127–141. (doi:10.1002/oa.594)
55. He C. 2018 Strategic analysis on the utilization of lithic raw material from Yumidong site in Wushan, Chongqing. *Quat. Sci.* **38**, 1449–1461.
56. He C. 2019 A preliminary research on the bone, antler and tooth artifacts from the Yumidong site in Chongqing. *Acta Anthropol. Sin.* **38**, 33–49.
57. Welker F, Smith GM, Hutson JM, Kindler L, Garcia-Moreno A, Villaluenga A, Turner E, Gaudzinski-Windheuser S. 2017 Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ* **5**, e3033. (doi:10.7717/peerj.3033)
58. Chen S, Pang L, Wu Y, Hu X. 2020 An assemblage of *Stegodon orientalis* fossils from the Yumidong Cave site in Wushan, Chongqing, with emphasis on the taphonomic analysis. *Chin. Sci. Bull.* **66**, 1482–1491. (doi:10.1360/TB-2020-0674)
59. Tong H, Patou-Mathis M. 2003 Mammoth and other proboscideans in China during the Late Pleistocene. *Deinsea* **9**, 421–428.
60. Ma J, Wang Y, Baryshnikov GF, Drucker DG, Mcgrath K, Zhang H, Bocherens H, Hu Y. 2021 The *Mammuthus-Coelodonta* faunal complex at its southeastern limit: a biogeochemical paleoecology investigation in Northeast Asia. *Quat. Int.* **591**, 93–106. (doi:10.1016/j.quaint.2020.12.024)
61. Tong HW, Li H, Xie JY. 2008 Revisions of some taxa of the Salawusu fauna from Sjara-Osso-Gol area, Neimongol. *China Quaternary Sci.* **28**, 1106–1113.
62. Boger U, Starkovich BM, Conard NJ. 2014 New insights gained from the faunal material recovered during the latest excavations at Vogelherd Cave. *Mitteilungen der Gesellschaft für Urgeschichte* **23**, 57–81.
63. Tong H, Wang X, Chen X. 2013 Late Pleistocene *Bison priscus* from Dabusu in Qian'an County, Jilin, China. *Acta Anthropol. Sin.* **32**, 485.
64. Norton CJ, Gao X. 2008 Hominin–carnivore interactions during the Chinese Early Paleolithic: taphonomic perspectives from Xujiayao. *J. Hum. Evol.* **55**, 164–178. (doi:10.1016/j.jhevol.2008.02.006)
65. Niven L. 2007 From carcass to cave: large mammal exploitation during the Aurignacian at Vogelherd, Germany. *J. Hum. Evol.* **53**, 362–382. (doi:10.1016/j.jhevol.2007.05.006)
66. Marān J *et al.* 2020 Neanderthal faunal exploitation and settlement dynamics at the Abri du Maras, level 5 (south-eastern France). *Quat. Sci. Rev.* **243**, 106472. (doi:10.1016/j.quascirev.2020.106472)
67. Titov VV. 2008 Habitat conditions for *Camelus knoblochi* and factors in its extinction. *Quat. Int* **179**, 120–125. (doi:10.1016/j.quaint.2007.10.022)
68. Andersen KK, Svensson A, Johnsen SJ. 2006 The Greenland ice core chronology 2005, 15–42 ka. Part 1: constructing the time scale. *Quat. Sci. Rev* **25**, 3246–3257. (doi:10.1016/j.quascirev.2006.08.002)
69. Rasmussen SO *et al.* 2014 A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. *Quat. Sci. Rev.* **106**, 14–28. (doi:10.1016/j.quascirev.2014.09.007)
70. Klementiev AM, Khatsenovich AM, Tserendagva Y, Rybin EP, Bazargur D, Marchenko DV, Gunchinsuren B, Derevianko AP, Olsen JW. 2022 First documented *Camelus knoblochi* Nehring (1901) and fossil *Camelus ferus* Przewalski (1878) from Late Pleistocene archaeological contexts in Mongolia. *Front. Earth Sci.* **10**, 861163. (doi:10.3389/feart.2022.861163)
71. Gautier A. 1993 The Middle Paleolithic archaeofaunas from Bir Tarfawi (Western Desert, Egypt). In *Egypt during the Last Interglacial: the Middle Paleolithic of Bir Tarfawi and Bir Sahara East* (eds F Wendorf, R Schild, AE Close), pp. 121–143. New York, NY: Springer. (doi:10.1007/978-1-4615-2908-8_8)
72. Gautier A, Makowiecki D, Paner H, Van Neer W. 2012 Palaeolithic big game hunting at Hp766 in Wadi Umm Rahau, Northern Sudan. *J. Afr. Archaeol.* **10**, 165–174. (doi:10.3213/2191-5784-10222)
73. Savioz NR, Morel P. 2015 La faune de Nadaouiyeh Aïn Askar (Syrie centrale, Pléistocène moyen): aperçu et perspectives. *Revue de Paléobiologie, Genève* **10**, 31–35.
74. Zavala EI, Jacobs Z, Vernot B, Shunkov MV. 2021 Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. *Nature* **595**, 399–403. (doi:10.1038/s41586-021-03675-0)
75. Vishnyatsky LB. 2004 The Middle-Upper Paleolithic interface in former Soviet Central Asia. In *The Early Upper Paleolithic beyond Western Europe* (eds PJ Brantingham, SL Kuhn, KW Kerry), pp. 151–161. Berkeley, CA: University of California Press.
76. Germonpré M, Lbova L. 1996 Mammalian remains from the Upper Palaeolithic site of Kamenka, Buryatia (Siberia). *J. Archaeol. Sci.* **23**, 35–57. (doi:10.1006/jasc.1996.0004)
77. Janz L, Rosen AM, Bukhchuluun D, Odsuren D. 2021 Zaraa Uul: an archaeological record of Pleistocene-Holocene palaeoecology in the Gobi Desert. *PLos ONE* **16**, e0249848. (doi:10.1371/journal.pone.0249848)
78. Khatsenovich AM *et al.* 2022 Shelter in an extreme environment: the Pleistocene occupation of Tsagaan Agui Cave in the Gobi Desert. *Antiquity* **96**, 1–9. (doi:10.15184/aqy.2022.51)
79. Tang Z, Liu S, Lin Z, Liu H. 2003 The late Pleistocene fauna from Dabusu of Qian'an in Jilin Province. *Vertebrata Palasiatica* **41**, 137.
80. Dong W, Hou Y-M, Yang Z-M, Zhang L-M, Zhang S-Q, Liu Y. 2014 Late Pleistocene mammalian fauna from Wulanmulan Paleolithic Site, Nei Mongol, China. *Quat. Int.* **347**, 139–147. (doi:10.1016/j.quaint.2014.05.051)
81. Qi G. 1975 Quaternary mammalian fossils from Salawusu river district, Nei Mongol. *Vertebrata Palasiatica* **13**, 239–249.
82. Griggo C. 2004 Mousterian fauna from Dederiyeh Cave and comparisons with fauna from Umm El Tlel and Douara Cave. *Paléorient* **30**, 149–162. (doi:10.3406/paleo.2004.4776)
83. Lyakhnitsky YS. 2018 Conservation of paleolithic cave art in the Capova cave. *J. Hist. Archaeol. Anthropol. Sci.* **3**, 137–141. (doi:10.15406/jhaas.2018.03.00075)
84. Devlet EG, Guillamet E, Pakhunov AS, Grigoriev NN, Gainullin DA. 2018 Preliminary results of studies of the camel figure at the Chamber of Chaos at Shulgan-Tash (Kapova) cave. *Ural Hist. J.* **58**, 141–148. (doi:10.30759/1728-9718-2018-1(58)-141-148)
85. Esin YN, Magail J, Monna F, Ozheredov YI. 2020 Images of camels on a mammoth tusk from West Siberia. *Archaeol. Res. Asia* **22**, 100180. (doi:10.1016/j.ara.2020.100180)
86. Wang N, Xu Y, Tang Z, He C, Hu X, Cui Y, Douka K. 2023 Large-scale application of palaeoproteomics (ZooMS) in two palaeolithic faunal assemblages from China. Figshare. (doi:10.6084/m9.figshare.c.6875422)

Supplementary Information for

Large-scale application of palaeoproteomics (ZooMS) in two Palaeolithic faunal assemblages from China

Naihui Wang*, Yang Xu, Zhuowei Tang, Cunding He, Xin Hu, Yinqiu Cui* and Katerina Douka*

Email: nwang@shh.mpg.de, cuiyq@jlu.edu.cn and katerina.douka@univie.ac.at

This SI file includes:

Supplementary text
Figures S1 to S9
Legends for Datasets Table S1 to S5
SI References

Other supplementary materials for this manuscript include the following:

Supplementary Table S1-S5
ZooMS spectra files (in both mzml and t2d formats) are currently accessible on Mendeley data ("ZooMS data of Yumidong and Jinsitai caves", Mendeley Data, V1, doi: 10.17632/ssf27rywhh.1)

Supplementary Text

Further information on sites and samples

Yumidong Cave was excavated from 2012 to 2015 with the archaeologists exposing about 150 m^2 surface area. The 6-metre-deep stratigraphic sequence includes 18 distinct layers, and the bedrock is not yet reached (Fig. S1). Currently, 48 dates have been obtained that range from the Middle Pleistocene to the Holocene, stretching the cave's sequence back to at least 230 ka (Layer 15). Bayesian analysis of 48 ages assigned layers 2 to 4 between 192-14 ka BP (95.4% confidence). A possible sedimentary discontinuity between 60-100 ka BP, marks a gap between layer 2 and layer 3. Layers 10 to 12 fall in 157-274 ka BP [1].

Jinsitai Cave is located in Inner Mongolia in northern China. The granite cave lies 1401 m above sea level, it was first excavated between 2000-2001 with the excavations reaching 5m below datum, Levallois artefacts were found from layer 5 and 6, which refers to the middle occupation phrase at Jinsitai. Ten years later, in 2012 and 2013, the second excavation campaign was addressed. Although only finite column deposit was left to dig, archaeologists recognized more Mousterian-like artefacts from the lowest layers at Jinsitai, rather than the middle layers [2]. The radiocarbon dates from the same layer suggest Mousterian-like industry appeared at least 47-42 ka, making Jinsitai an important site for exploring hominin dispersal.

ZooMS acid-insoluble protocol and MALDI-TOF analysis

Collagen was extracted and purified following established ZooMS protocols [22,28]. We used the acid-insoluble protocol. Samples were demineralised in 500 µL 0.5 M hydrochloric acid (HCl) for 24-48 h at 4 °C until the bone chips became spongy and reaction stopped. The acid supernatant was removed and the chips were then rinsed 3 times using 0.5 M ammonium bicarbonate (NH_4HCO_3) until a neutral pH was reached. The remaining sample was incubated at 65 °C for 1 h, in 100 µL of 50 mM NH_4HCO_3. Following incubation, 50 µL of the solution was collected and digested with 0.5 µg trypsin (Promega) at 37 ℃ for 18 h. Finally, 1 µL 5% trifluoroacetic acid (TFA) was added to stop the digestion. The digested samples were concentrated and desalted using C18 ZipTips (Thermo Scientific), then washed with 200 µL 0.1% TFA and eluted with 50 µL 50% acetonitrile / 0.1% TFA (*v/v*).

For peptide mass analysis, 0.5 µL of the elution was spotted with an equal volume of α -cyano-4-hydroxycinnamic acid solution (10 mg/mL in 50% acetonitrile / 0.1% TFA (*v/v*)) on an Opti-TOF 384

MALDI plate (AB SCIEX). One blank was analysed alongside every 23 samples as a negative control. Each sample or blank was spotted in triplicate. MALDI-TOF-MS analysis was carried out on a 5800 MALDI-TOF/TOF mass spectrometer (AB SCIEX) coupled with a 355 nm Nd-YAG laser based at Jilin University, China. The laser energy was adjusted to 4000 and all MS spectra were acquired using the reflector detection in the positive ionisation mode, collected in the *m/z* range of 800-3500. Calibration was performed using calibration mixture (AB SCIEX) to ensure mass accuracy within *m/z* 0.1.

Radiocarbon dating

Seven camel bones from Jinsitai were selected for dating and were dated at the Oxford Radiocarbon Accelerator Unit, University of Oxford, U.K. The bones were prepared for AMS dating using routine ultrafiltration methodologies as described in [38]. Briefly, about 600mg of coarsely ground bone powder was demineralised in 0.5 M HCl (18 h at room temperature (RT)), followed by 0.1M sodium hydroxide (NaOH) (30 min, at RT) and 0.5M HCl (~1 h, at RT) solutions, interspersed with ultra-pure (MilliQ™) water rinses between each reagent. The collagen was gelatinised in a pH 3 water solution at 75 °C for 20 h and filtered through previously-cleaned 9 mL polyethylene Ezee-filters™ (Elkay, UK). The filtrate was transferred by pipetting into previously-cleaned ultrafilters (Sartorius Vivaspin™ 15–30 kD MWCO) and centrifuged at 2500-3000 rpm until 0.5–1.0 mL of the >30 kD gelatin fraction was left (~ 30-40 min). The gelatin solution was collected using glass pipettes, placed into clean glass tubes, frozen at -18 °C, and dried for a minimum of 12 h. About 5mg of gelatin samples were combusted and analysed using a PDZ-Europa Robo-Prep combustion elemental analyser coupled to a PDZ-Europa 20/20 mass spectrometer operating in continuous flow mode using an He carrier gas. This enables $\delta^{15}N$ and $\delta^{13}C$, nitrogen and carbon content and calculation of C: N atomic ratios. Graphite was produced by reacting the sample CO_2 over an iron catalyst in an excess hydrogen atmosphere at 560 °C. AMS radiocarbon measurement was carried out using a Micadas accelerator.

Details on ZooMS data analysis

We performed semi-automatic identification with an own-built peak matching software followed by manual check of the identified peptides.

If a spectrum includes four or more markers out of nine matched with a species present in the ZooMS reference library, we regard the spectrum identifiable, or IDable. If less than four markers are found, the sample fails ZooMS identification. Three “unknown” samples found in this study, which means that while they had good collagen preservation, their markers did not match any taxa in the current ZooMS reference library [3–8].

Further to this, confident separation of key taxa is not always possible. For some samples, preservation of collagen was not sufficient to allow us to reach a genus/species level, especially for the Pecora infraorder, members of which usually share several markers. When 5 markers are identified (COL1α1 508-519/marker P1 for m/z 1105.6, COL1α2 978-990/marker A for m/z 1180.7, COL1α2 484-498/marker B for m/z 1427.7, COL1α2 793-816/marker D for m/z 2131.1 and COL1α1 586-618/marker F for m/z 2883.4) (JST 112 is an exception, which also has COL1α2 502–519/marker C for m/z 1550.6 and COL1α2 757–789/marker G for m/z 3059) the bone can belong either to a cervid or a bovid. Such peptide marker combinations are found to 47 IDable bones (7%) from Jinsitai Cave and in 1 from Yumidong Cave. These are assigned to Cervid/Bovid, but these samples do not belong to *Bos/Bison* or *Bubalus*.

The following ZooMS marker combinations at Jinsitai could represent three or four different Pecora taxa.

1) We assign 9 bones to “Antilopinae_1”. For these samples, COL1α2 978–990 (marker A, 1150 m/z) that is unique to genus *Rangifer* (reindeer) is identified, but COL1α2 757–789 (marker G, 3077 m/z) matches to the unique marker G for *Rangifer tarandus* and *Capra hircus* (capra). While reindeers are unlikely to exist at Jinsitai, we cannot therefore be certain which taxon these bones belong to. These fossils might be species with close ancestry to reindeer or capra that are not included in the current ZooMS library. Antilopinae_1 has 7 markers: COL1α1 508-519/marker P1 for m/z 1105.6, COL1α2 978–990/marker A for m/z 1150.7, COL1α2 484–498/marker B for m/z 1427.7, COL1α2 502–519/marker C for m/z 1550.6, COL1α2 793-816/marker D for m/z 2131.1, COL1α1 586-618/marker F for m/z 2883.4, COL1α2 757–789/marker G for m/z 3077.
2) There are 19 specimens with a marker combination we call “Antilopinae_2”. The species have the same markers included in *Moschus berezovskii*, *Pudu pudu*, *Rupicapra rupicapra* and *Ovis aries* all of which are present in the current reference library. The 7 markers for Antilopinae_2 are COL1α1 508-519/marker P1 for m/z 1105.6, COL1α2 978-990 marker A for m/z 1180.7, COL1α2 484–498/marker B for m/z 1427.7, COL1α2 502–519/marker C for 1580m/z, COL1α2 793-816/marker D for m/z 2131.1 and COL1α1 586-618/marker F for m/z

2883.4, with or without COL1α2 757–789/marker G for m/z 3017.

3) The Antilopinae_3 group includes possibly COL1α2 978–990 (marker A, 1192m/z) that is unique to *Ovibos moschatus* (musk ox), but COL1α2 502-519 (marker C, 1550m/z) does not match with the equivalent *Ovibos* marker. There is only one specimen at Jinsitai (JST 022) with such a marker combination, which is not included in the current ZooMS library.
The Antilopinae_3 group has 7 markers: COL1α1 508-519/marker P1 for m/z 1105.6, COL1α2 978–990/marker A for m/z 1193.7, COL1α2 484–498/marker B for m/z 1427.7, COL1α2 502–519/marker C for 1550m/z, COL1α2 793-816/marker D for m/z 2131.1 and COL1α1 586-618/marker F for m/z 2883.4, COL1α2 757–789/marker G for m/z 3017.
4) There are 53 specimens from Jinsitai and 4 from Yumidong with a marker combination. We assign it to “Cervidae/ Antilopinae”. The morphologically identified *Cervus elaphus* (red deer) at Jinsitai have this marker combination, together with many other species in the reference library. This group has 7 markers: COL1α1 508-519/marker P1 for m/z 1105.6, COL1α2 978–990/marker A for m/z 1180.7, COL1α2 484–498/marker B for m/z 1427.7, COL1α2 502–519/marker C for 1550m/z, COL1α2 793-816/marker D for m/z 2131.1 and COL1α1 586-618/marker F for m/z 2883.4, COL1α2 757–789/marker G for m/z 3017.

Stegodon is also not present in the current ZooMS reference library. Since the Proboscidea samples from Yumidong Cave share all available markers with that of five Elephantidae species in the reference library, the proboscideans from this site are assigned to Elephantoidae, a category that includes both Stegodontidae and Elephantidae.
There are two perissodactyls, *Stephanorhinus* sp. and *Tapirus terrestris* (tapir), in the current reference library; both share all diagnostic peptide markers. Unlike morphological approaches, ZooMS is not able to differentiate these genera. This limitation is reflected in the mixed category “Ceratomorpha" which includes extinct species *Megatapirus augustus* (giant tapir) and *Stephanorhinus kirchbergensis* (extinct two-horn rhino) at Yumidong Cave.
In the Jinsitai morphological bone assemblage, rodents, particularly bobak marmot, make about 14% of the entire assemblage. However, we were not able to identify marmots using ZooMS since *Marmota bobak* is not included in the reference library. In total, we find 4% (n=28) samples sharing most ZooMS markers with those known in the alpine marmot (CDS: XP_015350976.1). We tentatively assigned these Jinsitai samples to Rodentia and assume they probably belong to *Marmota bobak*.
Limited by the highly conserved sequence of COL1a1 and COL1a2 and the incomplete reference library for the region under study, the achieved taxonomic resolution on material from East Asia is much lower when compared to the morphological approach. The only exception when ZooMS outperforms the morphological approach is on differentiating *Bos/Bison* to *Bubalus sp.* as well as in the clear identification of camelids in the assemblage.

Additional statistical details

Sample weight significance test of Jinsitai Cave
Given the lack of stratigraphic provenance information, we grouped the Jinsitai samples by weight: <3g (n=248), 3-10g (n=228) and >10g (n=269) (Fig. S8). The identifiable rate is 88% (218/248), 91% (207/228) and 93% (249/269), respectively. To test if the identifiable rate of the >10g group is significantly higher than that of group <3g, Normal Approximation to the Binomial Distribution was employed. The result shows that the >10g group has a significantly higher IDable rate than group <3g ($z=2.30$, $p<0.05$).

Kernel Density Estimate in Figure 2 (insert)
Kernel Density Estimate is done with the R function geom_density() in package ggplot2 [9]. The parameters are defaulted as follows: stat_density(...,bw = "nrd0",adjust = 1,kernel = "gaussian",n = 512,…)

Fig. S1.

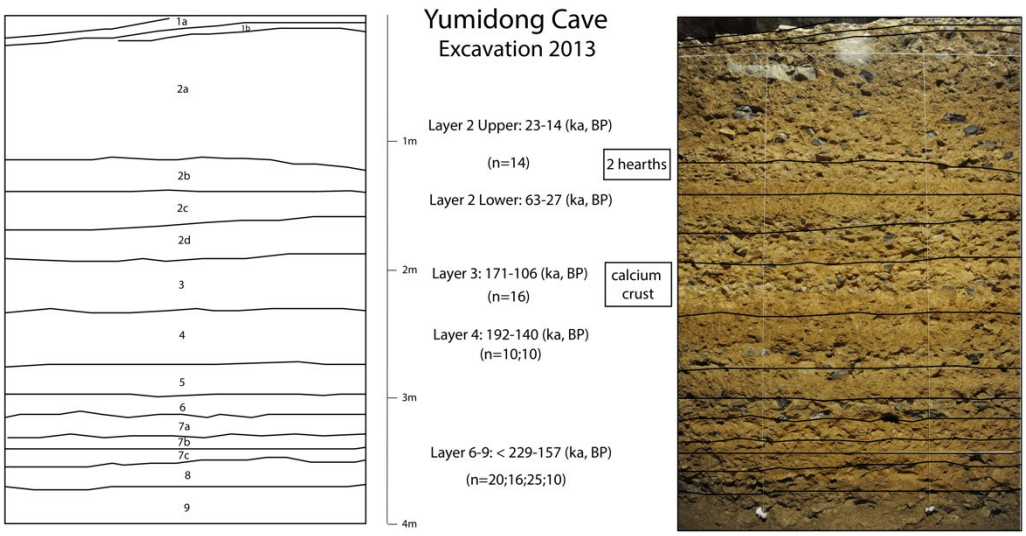


Figure S1. Stratigraphic layers of the east profile of Yumidong Cave 2013 excavation. In this study, 121 samples from 8 layers were analysed (layer 2, n=14; layer 3, n=16; layer 4, n=10; layer 5, n=10; layer 6, n=20; layer7, n=16; layer8, n=25; layer 9, n=10). Chronological data from [1].

Fig. S2.

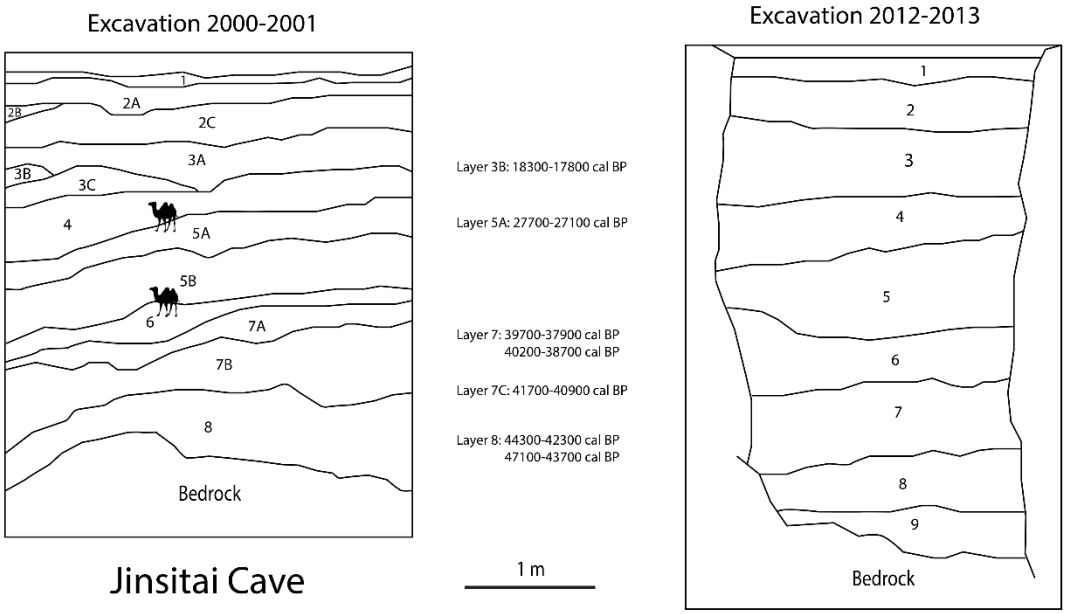


Figure S2. Stratigraphic profiles and radiocarbon dates of Jinsitai Cave, from [2,10]. The camels dated in this study fall into layer 3C to layer 6.

Fig. S3.

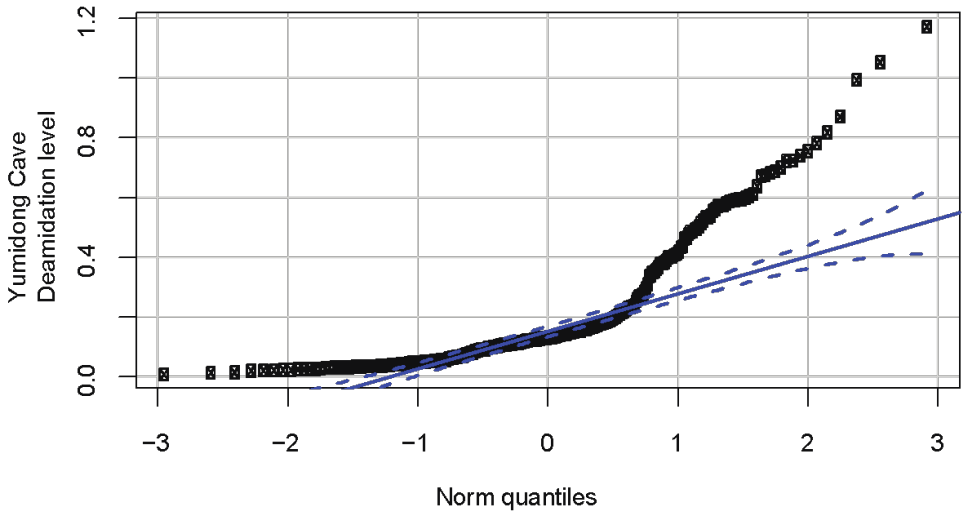


Figure S3. Normality test of Yumidong Cave deamidation rate (quantile-quantile plot), n=100.

Fig. S4.

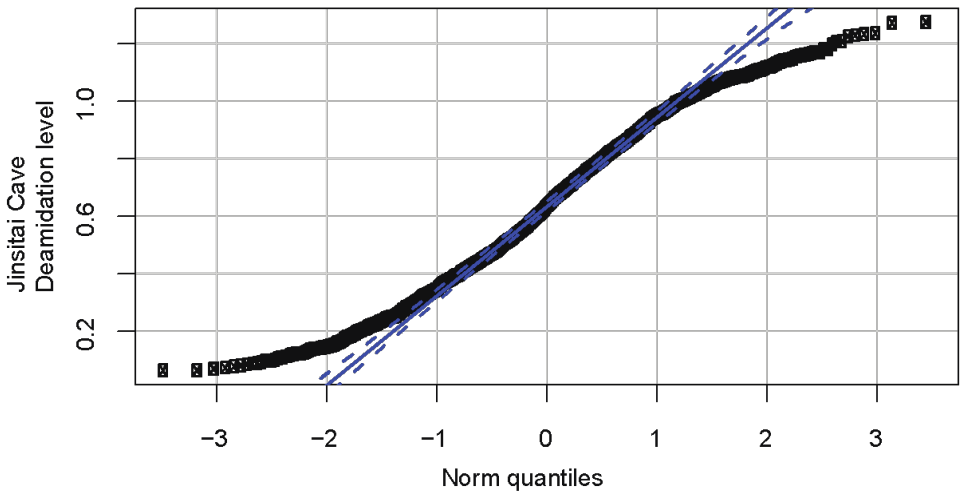


Figure S4. Normality test of Jinsitai Cave deamidation rate (quantile-quantile plot), n=676.

Fig. S5.

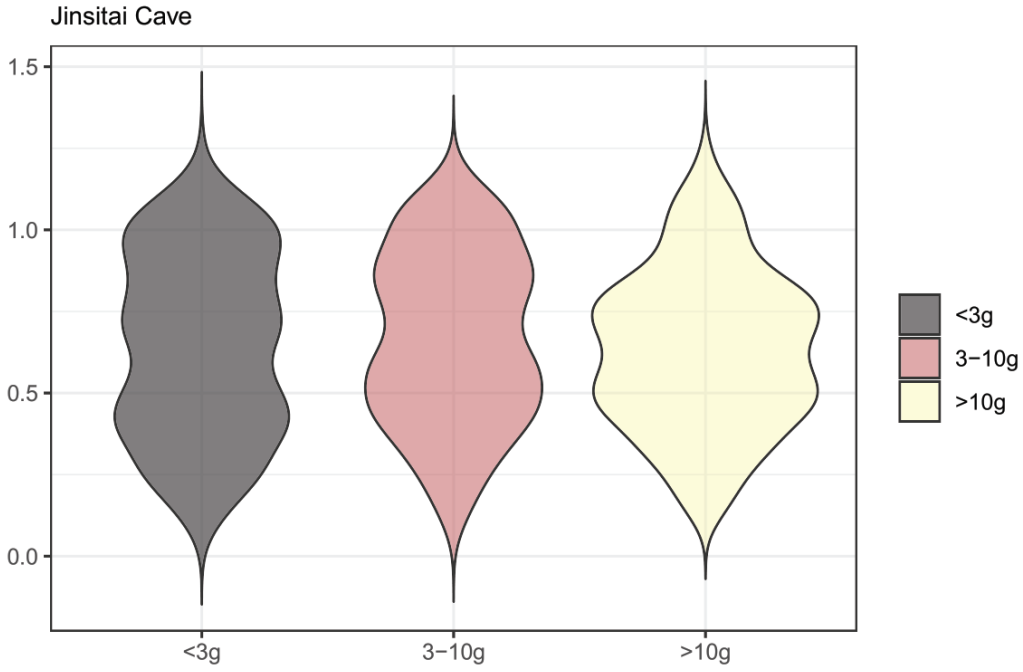


Figure S5. Violin distribution plot of Jinsitai Cave deamidation rate by weight. Weight <3g, n=218; 3-10g, n=207; weight >10g, n=249.

Fig. S6.

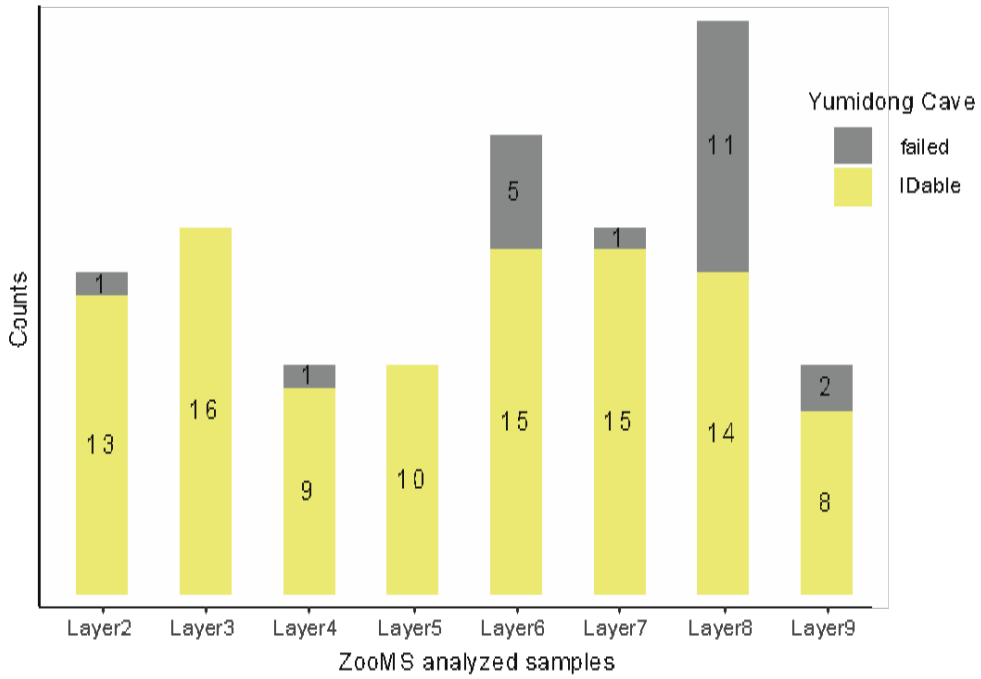


Figure S6. ZooMS-based identifiable results by horizontal layers, Yumidong Cave. Layer 2, n=14; layer 3, n=16; layer 4, n=10; layer 5, n=10; layer 6, n=20; layer7, n=16; layer8, n=25; layer 9, n=10.

Fig. S7.

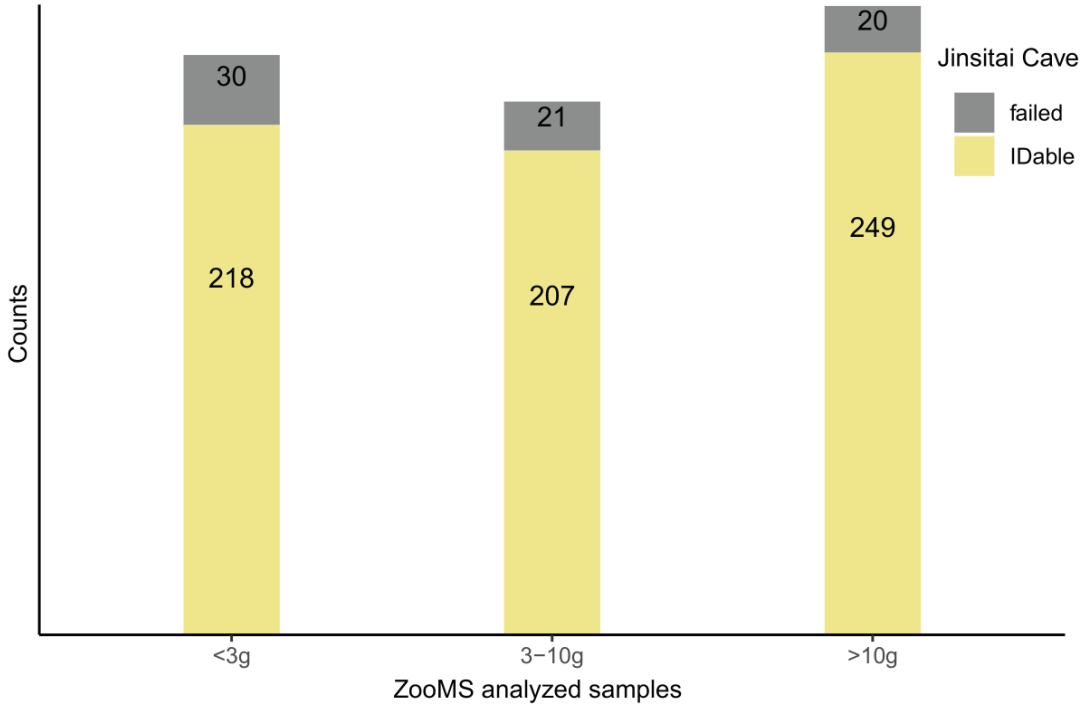


Figure S7. ZooMS-based identifiable results by weight groups, Jinsitai Cave. Weight <3g, n=248; 3-10g, n=228; weight >10g, n=269.

Fig. S8.

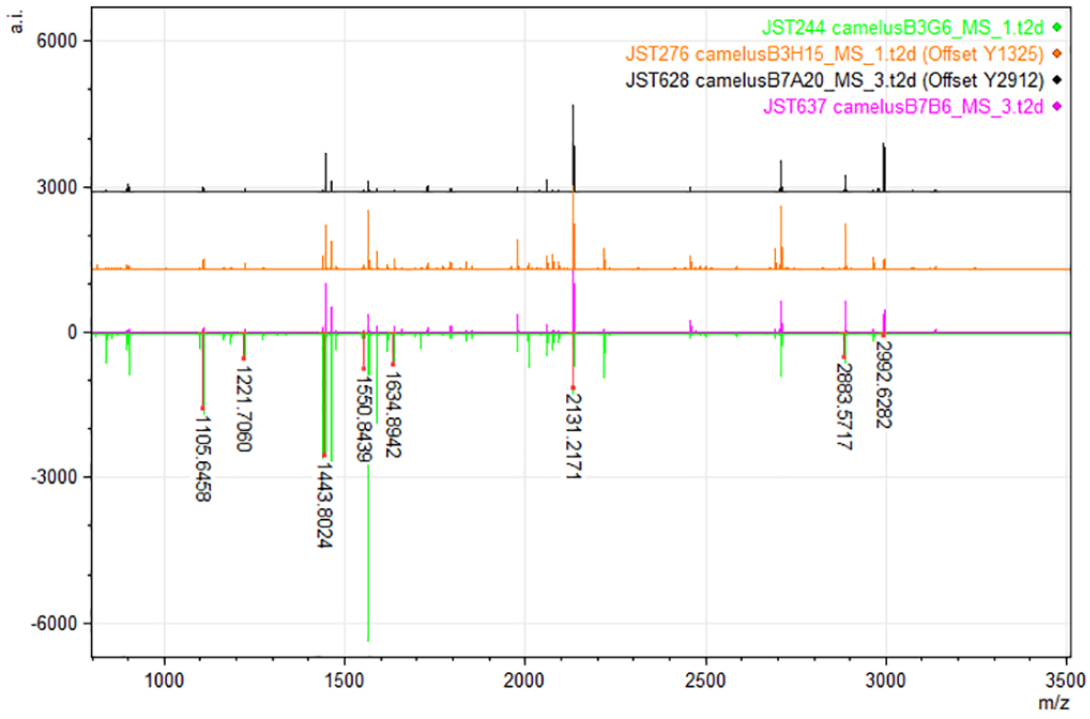


Figure S8. MALDI-ToF mass spectra of digested collagen from JST244 (green), JST276 (orange), JST628 (black) and JST637 (pink). Based on these spectra, these four bones were identified as *Camelus* sp..

Fig. S9.

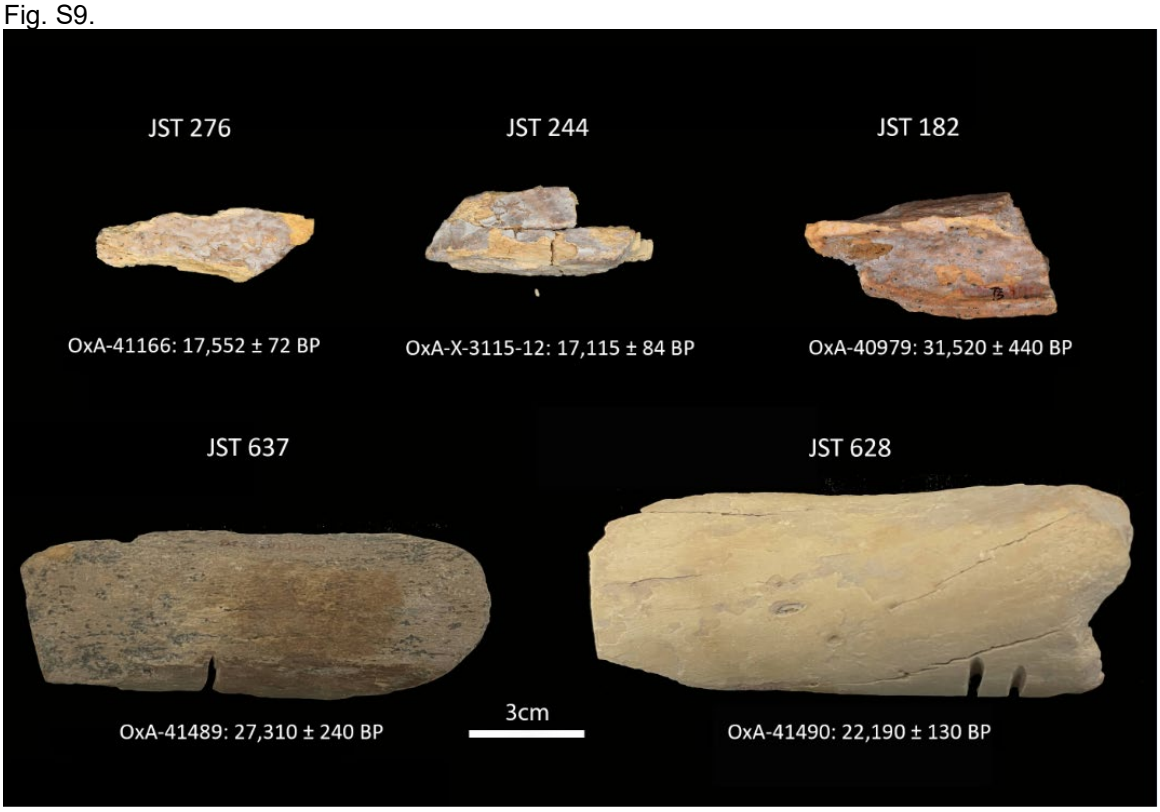


Figure S9. Five directly radiocarbon-dated Camelus sp. bones, and the uncalibrated radiocarbon ages. (More details in Table S5)

Legends for Datasets Table S1 to S5 (separate file)

Table S1. Deamidation of peptide COL1α1 508-519 or marker P1 (GVQGPPGPAGPR) of samples from Yumidong Cave (YMD) and Jinsitai Cave (JST). Each sample has three spectra except test samples JST001-JST007. For more details on the deamidation calculation see main text.

Table S2. Morphological and ZooMS identifications for Yumidong Cave and Jinsitai Cave faunal assemblages. The morphological datasets are from [11,12].

Table S3. Identification of the Yumidong Cave (YMD) faunal samples by ZooMS, based on collagen peptide markers, as described in REF [3–8] for mammals and birds. The nomenclature of ZooMS markers is standardised in [13].

Table S4. Identification of the Jinsitai Cave (JST) faunal samples by ZooMS, based on collagen peptide markers, as described in REF [3–8] for mammals and birds. The nomenclature of ZooMS markers is standardised in [13].

Table S5. New radiocarbon determinations of camel bones from Jinsitai calibrated using the IntCal20 curve [14]

SI References

1. Shao Q, Philippe A, He C, Jin M, Huang M, Jiao Y, Voinchet P, Lin M, Bahain J-J. 2022 Applying a Bayesian approach for refining the chronostratigraphy of the Yumidong site in the Three Gorges region, central China. *Quat. Geochronol.* **70**, 101304.

2. Li F *et al.* 2018 The easternmost Middle Paleolithic (Mousterian) from Jinsitai Cave, North China. *J. Hum. Evol.* **114**, 76–84.

3. 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 Commun. Mass Spectrom.* **23**, 3843–3854.

4. Buckley M, Kansa SW. 2011 Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. *Archaeol. Anthropol. Sci.* **3**, 271–280.

5. Buckley M, Larkin N, Collins M. 2011 Mammoth and Mastodon collagen sequences; survival and utility. *Geochim. Cosmochim. Acta* **75**, 2007–2016.

6. Rybczynski N, Gosse JC, Harington CR, Wogelius RA, Hidy AJ, Buckley M. 2013 Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Commun.* **4**, 1550.

7. Welker F *et al.* 2016 Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11162–11167.

8. Janzen A *et al.* 2021 Distinguishing African bovids using Zooarchaeology by Mass Spectrometry (ZooMS): New peptide markers and insights into Iron Age economies in Zambia. *PLoS One* **16**, e0251061.

9. Wickham H. 2016 ggplot2-Elegant Graphics for Data Analysis. Springer International Publishing. *Cham, Switzerland*

10. Wang X, Wei J, Chen Q, Tang Z, Wang C. 2010 A preliminary study on the excavation of the Jinsitai cave site. *Acta Anthropologica Sinica* **29**, 15–30.

11. He C. 2016 The lithic industry and human behaviour at Yumidong site. Jilin University. See https://cdmd.cnki.com.cn/Article/CDMD-10183-1017014181.htm.

12. Luo P. 2007 The research on the fauna and ancient ecological environment of the Late Pleistocene in Jinsitai cave site. Jilin University. See https://cdmd.cnki.com.cn/Article/CDMD-10183-2007105490.htm.

13. Brown S, Douka K, Collins MJ, Richter KK. 2021 On the standardization of ZooMS nomenclature. *J. Proteomics* **235**, 104041.

14. Reimer PJ *et al.* 2020 The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757.

Manuscript C

Naihui Wang, Nicholas J. Conard, Katerina Douka. “Integrating morphological and ZooMS-based approaches to zooarchaeology at Vogelherd Cave in Southwestern Germany”, *PaleoAnthropology*, Special Issue: Integrating ZooMS and Zooarchaeology: Methodological Challenges and Interpretive Potentials (2024).

Special Issue: Integrating ZooMS and Zooarchaeology: Methodological Challenges and Interpretive Potentials

Integrating Morphological and ZooMS-Based Approaches to Zooarchaeology at Vogelherd Cave in Southwestern Germany

NAIHUI WANG*
Max Planck Institute of Geoanthropology, 07745 Jena; and, Department of Early Prehistory and Quaternay Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen, GERMANY; nwang@gea.mpg.de

NICHOLAS J. CONARD
Department of Early Prehistory and Quaternay Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen; and, Senckenberg Centre for Human Evolution and Palaeoenvironment, Univesity of Tübingen, 72070 Tübingen, GERMANY; nicholas.conard@uni-tuebingen.de

KATERINA DOUKA*
Max Planck Institute of Geoanthropology, 07745 Jena, GERMANY; Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, 1030 Vienna; and, Human Evolution and Archaeological Sciences (HEAS), University of Vienna, 1030 Vienna, AUSTRIA; katerina.douka@univie.ac.at

*corresponding authors: Naihui Wang; *nwang@gea.mpg.de*; Katerina Douka; *katerina.douka@univie.ac.at*

submitted: 20 October 2023; revised: 3 March 2024; accepted: 4 March 2024

Guest Editors: Geoff M. Smith (School of Anthropology and Conservation, University of Kent, and Department of Archaeology, University of Reading), Karen Ruebens (Department of Archaeology, University of Reading, and Chaire de Paléoanthropologie, CIRB, Collège de France), Virginie Sinet-Mathiot (PACEA and Bordeaux Proteme-CBMN, Université de Bordeaux), and Frido Welker (Globe Institute, University of Copenhagen)

Handling Editor in Chief: Erella Hovers

ABSTRACT

Zooarchaeology is an established subfield of archaeology that incorporates a variety of interdisciplinary tools. Advances in analytical methods like radiocarbon dating, stable isotope analysis, and ancient DNA have added new dimensions to zooarchaeological research in the past century. In recent years, the addition of ZooMS (Zooarchaeology by Mass Spectrometry) has offered exciting new possibilities for studying faunal remains in archaeological contexts. In this study, we use the Vogelherd Cave, a Paleolithic site in the Swabian Jura of southwestern Germany, to showcase the advances in zooarchaeological analysis and changes in research focus. In 1931, G. Riek from the University of Tübingen completely excavated the site's rich deposits. In 2005–2012 and 2022–2023, N. J. Conard and a team from the University of Tübingen excavated Riek's backdirt using modern excavation techniques. The first systematic analysis of the faunal assemblage from a paleontological perspective was published by U. Lehmann in 1954, but it was not until the early 2000s that L. Niven undertook a comprehensive zooarchaeological study. In 2014, U. Boger and colleagues analyzed the faunal remains from the backdirt to gain a more complete view of the faunal assemblage. The current study adds the first ZooMS analysis at the site, focusing on 287 fragmentary bones obtained after water-screening the backdirt sediment. Here, we compile and compare our new ZooMS results to previous faunal datasets from Vogelherd. The history of research at the site provides a representative example of how the research focus has expanded over time and how novel analytical methods may contribute to the interpretation of an assemblage. Our ZooMS results represent the taxonomic abundance in a moderate way, which falls between the morphologically identified results of Niven and Boger et al. By juxtaposing traditional zooarchaeological results and ZooMS data, we explore the strengths and weaknesses of each approach and contemplate how best to integrate these methods in future research.

INTRODUCTION

Zooarchaeology is a multidisciplinary field that involves the study of animal remains from archaeological sites to reconstruct human-animal interactions in specific environments (Reitz and Wing 1999). This type of research initially originated alongside prehistoric archaeology. In the mid-19th century, faunal studies began to focus on domestic animals or modified bone tools, leading to cultural interpretations (Eaton 1898; Forchhammer 1852; Mills 1904). Typically, zoologists or paleontologists conducted these early faunal studies. In the 1950s, guidelines on animal bone sorting and identification for archaeologists became available (Cornwall 1956; Lawrence 1951). Zooarchaeological research evolved beyond species lists to a dynamic field investigating all aspects of past human-animal-environment interactions. In 1971, Olsen (1971) proposed the term "zooarchaeology'' to describe the study of animal remains to answer archaeological questions. Following this, zooarchaeological quantitative methods and report paradigms were developed and published (Brumley 1973; Grayson 1979; Grigson 2016). Discussions about the identification of cutmarks and fragmentation patterns on bones, such as those conducted by Behrensmeyer and colleagues (1986), Binford (1981) and Johnson (1985), have contributed to the methodology of zooarchaeology. Archaeology has also updated the fieldwork standards and more widely adopted practices, such as sediment screening (Geiling et al. 2018), that provided enlarged faunal assemblages for zooarchaeological studies.

The application of molecular analytical methods, such as isotopes, radiocarbon dating, and ancient DNA (aDNA), has enriched the zooarchaeological endeavor by providing direct information about age, diet, and phylogenetics. aDNA analysis is an effective tool for assigning morphologically tricky specimens and understanding the history of animal domestication (Horsburgh 2008; Librado et al. 2021). In 2009, a paleoproteomics method, ZooMS (Zooarchaeology by Mass Spectrometry), was introduced as a taxonomic identification tool (Buckley et al. 2009). Due to its low cost and potential for high throughput, ZooMS offers a sustainable solution to taxonomically identifying large numbers of fragmented bones from archaeological sites. ZooMS outperforms other taxonomic techniques focusing on fragmented bones, e.g., macroscopical observation (Cuijpers 2006) or metabarcoding DNA (Grealy et al. 2015), and provides extensive taxonomic data for zooarchaeology (Brown et al. 2021; Buckley et al. 2017; Martisius et al. 2022; Pothier-Bouchard et al. 2020; Ruebens et al. 2023; Sinet-Mathiot et al. 2023, 2019; Torres-Iglesias et al. 2024). However, concerns about the way of integrating traditional zooarchaeological results and ZooMS datasets are being raised (Banning 2020; Giovas and LeFebvre 2017).

In this study, we examine the evolutionary trajectory of the field of zooarchaeology through the lens of the Vogelherd Cave case. The archaeological studies at Vogelherd have produced a wealth of information regarding the prehistoric occupants and their choices, and the paleoecological framework they existed in. Starting with a brief excavation history of Vogelherd (Figure 1), we introduce the cultural attributions of the original deposit in the cave. Subsequently, we summarize the three traditional zooarchaeological studies of Vogelherd's faunal collections, in terms of methods, results, and representative conclusions they each reached. We then present our latest results based on the application of ZooMS on a small faunal sub-set of Vogelhed. Finally, we undertake a comparative analysis of the four zooarchaeological studies, delving into their potential for integration. The paper provides a historical overview of the development of zooarchaeology, since each dataset represents a typical research of its time.

VOGELHERD CAVE EXCAVATION HISTORY

Vogelherd Cave is part of the karst system in the Swabian Jura of southwestern Germany. Located in the Lone Valley, between the Danube River to the south and the Jurassic Plateau to the north, the cave, with its three entrances, covers an area of ca. 170m^2. This site provides outstanding views of the Lone Valley and has a comfortable scale, making it a desirable location for human habitation.

Gustav Riek from the University of Tübingen and a small team of local workers excavated the site in twelve weeks during the summer of 1931 and removed around 500m^3 of deposits. Riek documented twelve stratigraphic profiles and identified nine cultural horizons spanning the Neolithic to the Middle Paleolithic, which he published in his famous monograph in 1934.

The Archaeological Horizon I (AH), specifically assigned to the Neolithic period. AH II and III, deposited in the early phase post the Last Glacial Maximum, are characterized as the Magdalenian. The Magdalenian horizons at Vogelherd, as inferred from low find density, suggested short-time stays at Vogelherd (Niven 2006; Riek 1934).

Riek described no Gravettian layer at Vogelherd. Nevertheless, subsequent archaeological investigations identified potential Gravettian blades and points, and the radiocarbon date on a bone from AH IV fell within the Gravettian period (ca. 26 ka), suggesting that there was Gravettian component at Vogelherd (Conard and Bolus 2003; Conard et al. 2012). AH IV and V were referred to as the "upper and middle Aurignacian" by Riek (Riek 1934). Radiocarbon dates from the two Aurignacian horizons at Vogelherd, and corroborating evidence from other Swabian Jura sites firmly establish them as among the earliest Aurignacian technocomplexes in Europe, dating to ca. 43–35 ka (Conard and Bolus 2003; Conard et al. 2004; Higham et al. 2012).

Originally designated as the "lower Aurignacian," AH VI was latter re-assigned to the Middle Paleolithic (Müller-Beck 1957). AH VII and AH VIII were similarly attributed to the Middle Paleolithic by Riek. Excavating down to the bedrock, Riek assigned AH IX as the "culture of the cave floor" (Riek 1934). This horizon, composed of ochre-yellow loam and bean ore, likely dates back to Marine Isotope Stage 5e, a period of warmer climate, a conclusion supported by the recovery of a molar tooth from a forest elephant that thrived in such conditions (Niven 2006).

The Aurignacian horizons AH V and IV represent by

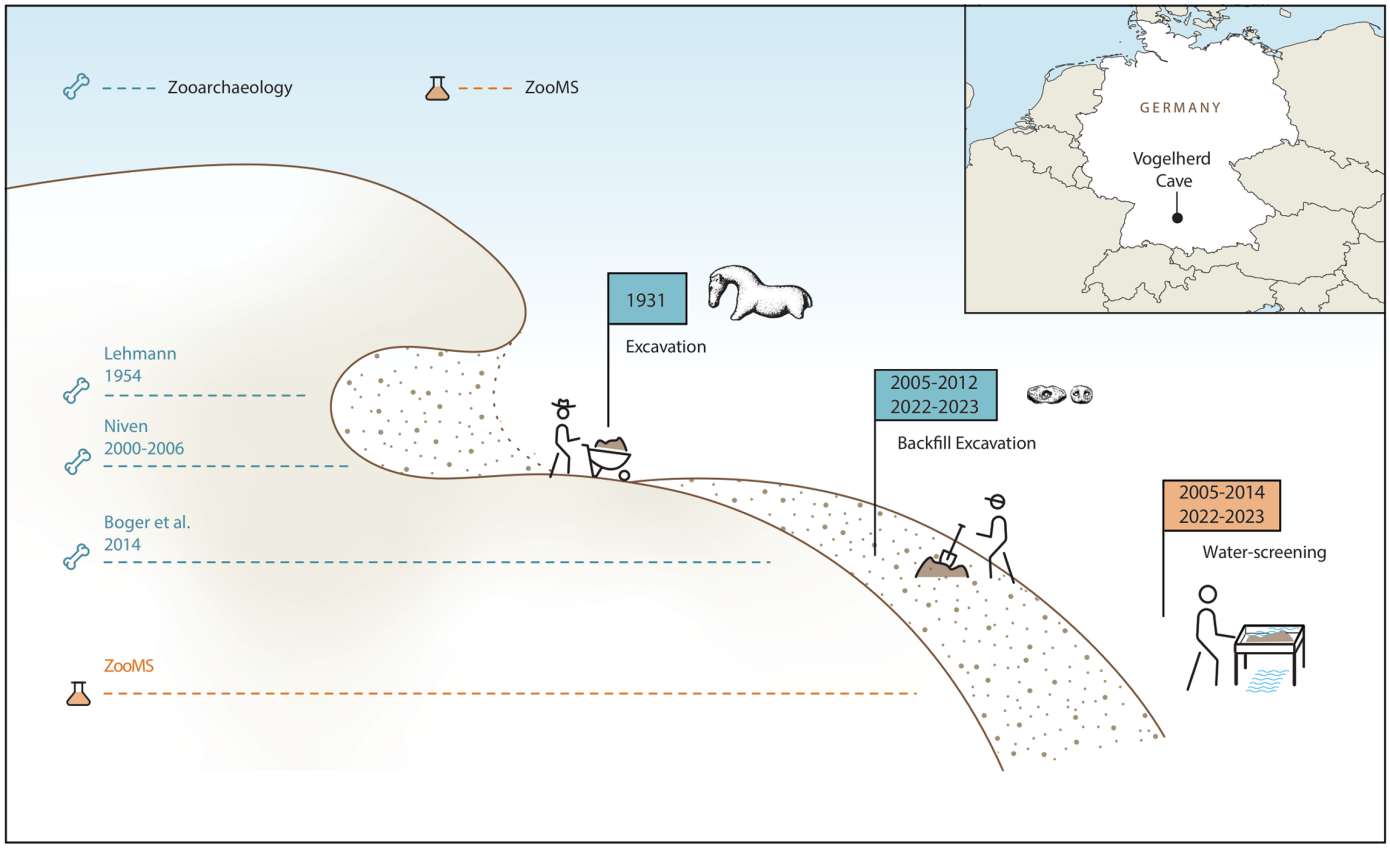


Figure 1. The schematic of Vogelherd Cave highlights two main excavations and the four faunal assemblage studies.

far the richest layers at Vogelherd and yielded 2,863 lithic tools and thousands of blanks, as well as nine ivory figurines and numerous other organic artifacts (Hahn 1977; Riek 1934). While the excavation from 1931 lacked piece-plotting, careful scrutiny of the documentation and labeled finds sometimes provides a degree of contextual information beyond the stratigraphic attribution, the refits of lithic and faunal material show a degree of mixing between the layers (Schürch in prep.).

In addition to the lithic and other artifact assemblages, Riek's excavation produced more than 18,000 bones, making this one of the largest Upper Paleolithic faunal assemblages in Central Europe (Niven 2007). Along with other human remains, Riek recovered a modern human skull, known as the Stetten 1 cranium, from the base of Aurignacian Horizon V. This fossil long served as evidence for assigning the site's figurative art to modern humans (Riek 1934). Direct AMS dating of the Stetten 1 cranium and other human skeletal material, however, revealed that they were of Neolithic age, approximately 5,000 years ago (Conard 2009; Conard et al. 2004).

In order to gain additional information and contextualize the results from Riek's dig, from 2005 to 2012 and 2022 to 2023, a team from the University of Tübingen under Nicholas J. Conard's direction re-excavated the backdirt from the original fieldwork on a slope outside Vogelherd. Conard's team excavated the entire volume of sediment from the cave bucket by bucket while piece-plotting numerous single finds and then water-screening all the sediments. As a result, the team recovered over 200,000 lithic artifacts, hundreds of Aurignacian ivory beads, numerous fragments of figurative artworks and musical instruments, and countless unidentified fragments of osseous material (Conard et al. 2015a).

ZOOARCHAEOLOGICAL STUDIES AT VOGELHERD CAVE

THE FIRST VOGELHERD FAUNAL ANALYSIS OF THE 1931 EXCAVATION

> *"The Vogelherd offers the researchers the great advantage as it was carefully excavated horizontally and well-documented."* (Lehmann 1954: 144)

Paleontologist Ulrich Lehmann conducted the first study of the 1931 Vogelherd faunal assemblage. In his publication, Lehmann appreciated the methods and recording of Riek's excavation.

Although some contextual information was available, the results were listed taxonomically in Lehmann's publication. Lehmann only examined intact fossils. He identified species and body sizes, based on teeth and the more complete bones that were identifiable and measurable. The intact fossils, however, constituted only a fraction of the

assemblage collected at the site (n=921). Morphology and morphometrics dominated Lehmann's study, reflecting his paleontological interest. He compared the morphometric data within Vogelherd using available data from other sites and modern collections (illustrated in tables and line charts). The statistical differences showed changing trends, and Lehmann suggested that the size of animals reflected an adaptation to paleoecological changes. Lehmann identified all 921 specimens to at least genus level, with even subspecies identification for equids (Table 1); he did not mention fragmentary and unidentifiable specimens. The paleontologist also excluded the Elephantidae remains, which comprised the most significant proportion of the Vogelherd faunal assemblage. These remains were dispatched to another Elephantidae paleontologist, Karl Dietrich Adam, whose work was never published. Lehmann only cited Elephantidae's minimum number of individuals (MNI) based on Adam's study. MNI has a more notable position than the number of identified specimens (NISP) in Lehmann's study, where he listed the MNI values of each taxon in a table but only included NISP values in the description text. Element counts were not fully reported, and how the MNI values were derived was also unclear.

At the end of the publication, Lehmann (1954) attempted to reconstruct the paleoclimate by comparing taxonomic abundances and body size differences among cultural horizons. He concluded that the Middle Paleolithic period at Vogelherd was cool and the Upper Paleolithic (Aurignacian and Magdalenian) was colder, due to the presence of more Arctic fauna; however, the climate at Vogelherd was warm prior to the occupation of hominin. This climate reconstruction on fauna was generally correct through later climate reconstruction research on other proxies (Andersen et al. 2006; Rasmussen et al. 2014). The last glacial maximum between Aurignacian and Magdalenian occupation left almost no fauna fossils at Vogelherd, making it hardly detectable for Lehmann. Regarding the formation of the Vogelherd deposit, Lehmann assumed that hominins had introduced the faunal remains to the cave, so it is not a carnivore den or natural trap. He concluded that horses, mammoths, and reindeer were the main prey targets of the occupants at Vogelherd, based on their high MNI values.

Unlike a species list (Lyman 2015a) that was common in his time, Lehmann's study is closer to what was later defined as zooarchaeology. He published his work on its own, not as an appendix in a monograph; he quantified and produced exact values of MNI and NISP, rather than describing them as "rare" or "common." Although modifications were beyond the scope of his paleontological research, he also observed and recorded a few modification traces as cutmarks on reindeer and wolf remains. Finally, he wrote two sections on interpreting human subsistence strategies, and the study ends with the relationship between ancient humans and their ecological context.

THE COMPLETE ZOOARCHAEOLOGICAL STUDY OF THE 1931 FAUNAL ASSEMBLAGE

> "*Initially, an attempt was made in this study to match specific finds described or depicted in Lehmann's report with those in the existing collections*." (Niven 2006: 7).

Between 1999 and 2004, zooarchaeologist Laura Niven studied the fauna from the 1931 excavation as a doctoral candidate in the German Science Foundation's Collaborative Research Centre 275: Climate-coupled processes in the Mesozoic and Cenozoic eras based at the University of Tübingen. Her work, which was supervised by Nicholas J. Conard and Hans-P. Uerpmann, represents the first comprehensive study of the faunal material from the Vogelherd Cave (Niven 2001, 2003, 2006, 2007). During her initial attempts to match the specimens with Lehmann's descriptions, Niven found it challenging to locate catalog numbers after nearly 50 years. At the time of Niven's research, the Vogelherd faunal collection was housed at four locations. The cataloged faunal remains comprised around 18,800 specimens, including osseous tools and artifacts. Among these, 14,181 specimens were >1cm long and preserved stratigraphic information. These 14,181 findings formed the central focus of Niven's study. According to Niven (2006), ca. 94% of the studied specimens (n=13,282) were from the Aurignacian horizons. However, the overlap with Lehmann's study was not exact, given loss, curation, breakage, and refits over 50 years. Lehmann's earlier study occasionally documented specimens in more complete preservation than later studies.

Niven's study assessed the completeness of the 1931 excavation by quantifying the types of long bone circumferences and determining the minimum number of elements (MNE) in the Aurignacian assemblage (Niven 2006). The results showed that, unlike other early excavations that often discarded long bone shafts (Marean 1998; Marean et al. 2004), bones in Vogelherd Aurignacian horizons were equally collected, except those <3cm in length. The collection completeness analysis laid a strong foundation for Niven's subsequent reconstruction of human behavior at the site.

Niven (2006) taxonomically determined 7,730 specimens to family, genus, or species (see Table 1). For the unidentified fragments, Niven classified them using broader designations such as "large artiodactyl" or by body sizes. In addition to Lehmann's taxa list, Niven described a few bird species (n=13) for the first time in the Aurignacian context. Mammoths accounted for 46% of Niven's identifiable assemblage. Niven suggested that Aurignacian humans collected large quantities of mammoth bones and ivory not for food, but for other particular uses. Thus, mammoths were not regarded as one of the main prey at Vogelherd (Niven 2001, 2006).

TABLE 1. TAXON TALLIES AND COMPOSITIONS OF ZooMS, ARCHAEOLOGICAL LAYERS AH II-IX BY LEHMANN (1954), NIVEN (2006), AND GEOLOGICAL UNIT HL/KS (Boger et al. 2014). (Note that the bold cells refer to ZooMS-only categories.)

				Morphological approach					
				AH II-IX (Lehmann 1954)		AH II-IX (Niven 2006)		HL/KS (Boger et al. 2014)	
Identification		ZooMS							
Categories	Taxon	counts	%	NISP	%	NISP	%	NISP	%
Aves	Birds					13	0.17	15	0.76
Hominidae	***Homo sapiens***	3	1.22						
Erinaceidae	*Erinaceus europaeus*							7	0.36
Rhinocerotidae	*Coelodonta antiquitatis*	12	4.90	73	7.93	185	2.39	33	1.68
Equidae	*Equus ferus*	45	18.37	586	63.63	1825	23.61	229	11.64
Leporidae	*Lepus* sp.	23	9.39	3	0.33	30	0.39	106	5.39
Ursidae	*Ursus* sp.	5	2.04	37	4.02	152	1.97	129	6.56
ungulate prey taxa	**Pecora**	18	7.35						
	Rangifer tarandus	28	11.43	115	12.49	1679	21.72	123	6.25
	Rangifer tarandus/Capra hircus	4	1.63						
	Bos/Bison	1	0.41	19	2.06	111	1.44	52	2.64
	Capra hircus	2	0.82						
	Capra hircus/Ovis aries							25	1.27
	Ovis aries/Rupicapra rupicapra	1	0.41						
	Cervus elaphus/Megaloceros giganteus	6	2.45						
	Cervus elaphus			15	1.63	29	0.38	11	0.56
	Megaloceros giganteus			10	1.09	8	0.10		
	Capreolus capreolus							80	4.07
	Rupicapra rupicapra			2	0.22	2	0.03		
	Ovibos moschatus			1	0.11				
	Sus scrofa	2	0.82	2	0.22	8	0.10	35	1.78
	Sub-total	62	25.31	164	17.81	1837	23.66	326	16.57

TABLE 1. TAXON TALLIES AND COMPOSITIONS OF ZooMS, ARCHAEOLOGICAL LAYERS AH II-IX BY LEHMANN (1954), NIVEN (2006), AND GEOLOGICAL UNIT HL/KS (Boger et al. 2014) (continued). (Note that the bold cells refer to ZooMS-only categories.)

				Morphological approach					
				AH II-IX (Lehmann 1954)		AH II-IX (Niven 2006)		HL/KS (Boger et al. 2014)	
Identification		ZooMS							
Categories	Taxon	counts	%	NISP	%	NISP	%	NISP	%
Elephantidae	*Elephas antiquus*			1	0.11	1	0.01		
	Mammuthus primigenius			excluded		3585	46.38	686	34.88
	Mammuthus primigenius/Elephas antiquus	79	32.24						
	Sub-total	79	32.24	1	0.11	3586	46.39	686	34.88
Canidae	*Canis lupus*			29	3.15	45	0.58	74	3.76
	Canis lupus/Vulpes lagopus	2	0.82						
	Vulpes vulpes	6	2.45	2	0.22				
	Vulpes lagopus			5	0.54				
	Vulpes sp.					20	0.26	102	5.19
	Sub-total	8	3.27	36	3.91	65	0.84	176	8.95
other carnivores	**Felidae/Hyaenidae/Mustelidae**	2	0.82						
	Mustelidae	4	1.63						
	Meles meles			3	0.33	excluded		188	9.56
	Gulo gulo			2	0.22	1	0.01	1	0.05
	Martes sp.							5	0.25
	Mustela sp.							1	0.05
	Panthera leo spelaea			4	0.43	6	0.08	2	0.10
	Crocuta crocuta spelaea			12	1.30	27	0.35	27	1.37
	Felis silvestris					3	0.04	36	1.83
	Sub-total	6	2.45	21	2.28	37	0.48	260	13.22
	unknown for ZooMS	2							
	failed identification	42	14.63	not available		5620	42.10	375	16.01
	Total	287		921		13350		2342	

Unlike Lehmann, Niven is a zooarchaeologist. This background difference is reflected in the organization of their publications. In the initial part of Niven's (2006) book, the author explained the zooarchaeological methods she would employ, including the definitions for quantification units and anthropogenic and non-anthropogenic modifications. She also briefly introduced the history and paleoecology of the mammal species found at Vogelherd. The author presented the study in a way most relevant to addressing archaeological interests, in chronological sequence—Middle Paleolithic, Aurignacian, and Magdalenian. She evaluated the properties of each cultural deposit, whether carnivores or hominins were the main accumulators, with multiple zooarchaeological proxies including standard quantification units (e.g., NISP, MIN, MNE, and MAU), modification, age, and sex profiles of the studied fauna. Niven also measured the morphometric data for establishing taxa age profile and season-at-death. In the last chapter, she places the Vogelherd faunal assemblage in the context of the Lone Valley and the Swabian Jura and summarizes the human subsistence behavior (Niven 2006).

THE PLOTTED FAUNAL ASSEMBLAGE FROM VOGELHERD BACKDIRT

> "*The excavated sediments were all screened, and further insights into small game exploitation will be revealed after analyzing these remains*" (Boger et al. 2014).

From 2005 to 2012 and 2022 to 2023, a team from the University of Tübingen under Nicholas J. Conard's direction re-excavated the backdirt from the original fieldwork on a slope outside Vogelherd Cave. The 2005–2012 dig cleared nearly 90% of the sediment previously in the cave. Conard's team excavated sediment by bucket unit, while piece-plotting numerous single finds that were visible in the field. Following the first re-excavation of the backdirt, Boger and colleagues 2014) analyzed these plotted faunal remains.

Given that reconstructing the history of a backdirt formation is not always feasible (e.g., Wright et al. 2021) and rarely reliable, Boger et al. (2014) examined the backdirt faunal assemblage (n=2,342) as a whole, regardless of the bone locations in the backdirt. As Niven (2006) claimed that 94% of the faunal remains in the 1931 excavation came from intact Aurignacian horizons, Boger and colleagues assumed a similar proportion of Aurignacian remains in the 2005–2012 handpicked assemblage. Moreover, they tried to assign the faunal remains to their original context (Middle Paleolithic, Aurignacian, or Magdalenian) based on the state of preservation. However, this was not possible because of the variability of bone preservation and differences between *in situ* finds and the backdirt assemblage (Boger et al. 2014).

Boger et al. (2014) identified nearly 84% of the handpicked assemblage to genus level or higher (see Table 1), representing a high identification rate for *in situ* Paleolithic horizons in the Lone Valley (Kitagawa 2014). For the unidentified specimens (n=375), Boger et al. (2014) assigned them to four body size groups following Brain's (1983) classification. Regarding faunal diversity, this study added a few new taxa to Vogelherd's previously known faunal list. These new taxa included small mammals such as roe deer, marten, polecat, and hedgehog (n=93). Intrusive animals, such as badgers (n=188), and domesticated animals like sheep/goats (n=25), were excluded in Niven's study, but included by Boger et al. (2014). These species, however, likely played little or no role in the subsistence of the Paleolithic inhabitants of Vogelherd Cave.

Regarding species abundance through NISP values, the most noticeable difference compared with the *in situ* assemblages was the higher percentage of small animals, such as wild hares, in the backdirt. Clearly, Riek's excavation overlooked much of the small and highly fragmented faunal material, which resulted in a less prominent presence of the two main prey species, horse and reindeer, in the backdirt assemblage. The composition of mammoths remained similar, but in the backdirt, mammoths were notably over-represented by ivory fragments, which accounted for 94% of the total NISP for mammoths. Retrieved from the backdirt, the generally smaller size of specimens attributed most of the differences observed in the comparison.

Results in Boger et al. (2014) reflected a heightened interest in the potential role of the small and low-ranked game, especially wild hares (n=106), with emphasis on examining diet breadth and resource ranking during the Paleolithic. They also argued that water-screened sediments would reveal evidence of a higher level of small animals (Boger et al. 2014). This prediction becomes testable when we apply ZooMS to the highly fragmented faunal remains recovered during water screening.

ZOOMS-ANALYZED WATER-SCREENED BONE FRAGMENTS FROM VOGELHERD

ZooMS is a paleoproteomics approach providing a low-cost, fast, and reliable way of speciating collagen materials such as bone, ivory, and leather. Collagen is the major organic component (~90%) in the bone of vertebrates, and its survival was tested in many Paleolithic sites (Richter et al. 2022). ZooMS involves extracting Type I collagen (COL1) from a sample and generating tryptic-digested peptide mass fingerprints via MALDI-TOF-MS (Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry).

Sediments excavated in 2005–2012, collected from over 32,000 buckets, were water-screened using a 2mm sieve. This work was completed by 2014, two years after the excavations concluded (Conard et al. 2015a). By sorting the dry sediment, identifiable microfauna bones, teeth/ivory fragments, burnt bones, and tiny artifacts like ivory beads were separated from highly fragmented bones. The sorting resulted in >100kg of bone fragments, stored in bags and recorded with bucket units (Conard et al. 2015b; Schuerch et al. 2021). Conard's team found hundreds of ivory beads in the backdirt, echoing the discovery of an ivory rod cache in the 1931 excavation (Riek 1934), likely for bead-making. Most beads have similar characteristics to those from other

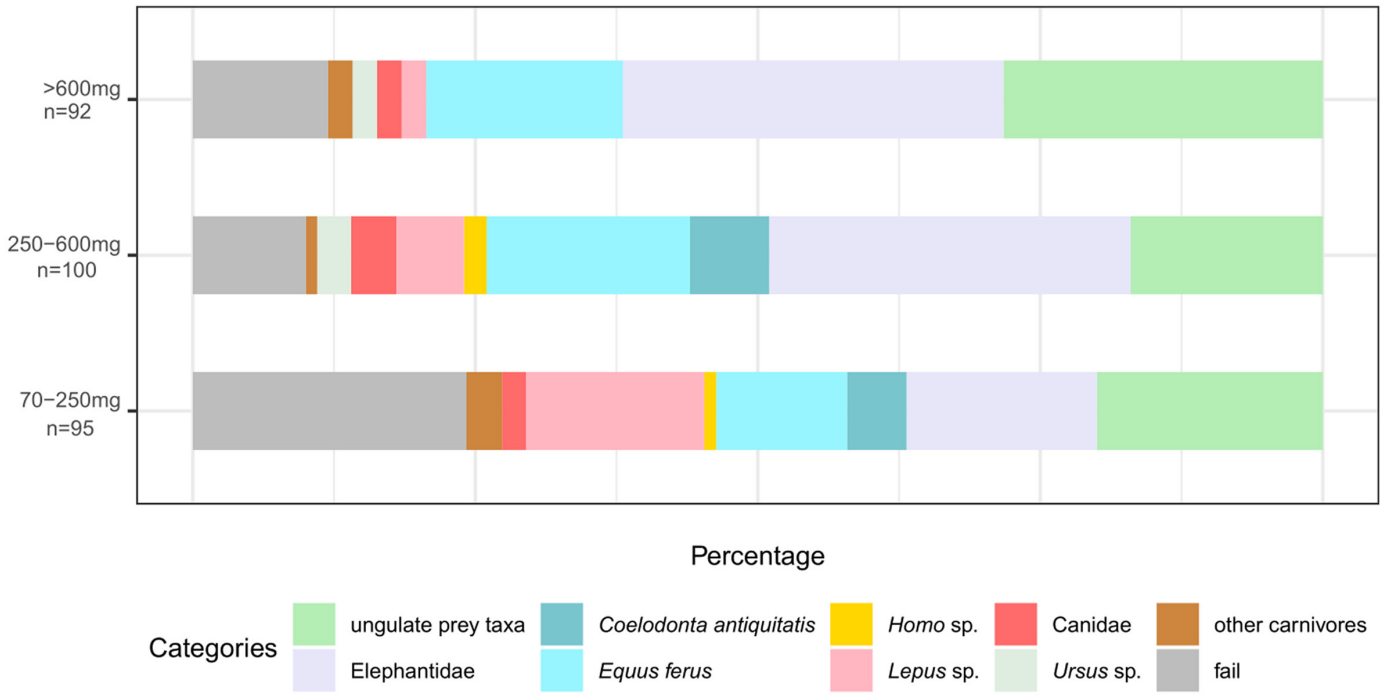


Figure 2. ZooMS identifications grouped in weights. Group 70–250mg, n=95; Group 250–600mg, n=100; Group >600mg, n=92.

Swabian Jura caves dating to the Aurignacian period (Hahn 1988; Wolf and Conard 2015).

To apply ZooMS in this assemblage, we selected random bags of bones from buckets containing ivory beads. The sampled material came from the front of the southwestern entrance of Vogelherd Cave (Supplementary Material Figure S1). We sampled bones with initial weight >70mg for ZooMS (n=287), since smaller bones would preclude subsequent analyses, such as radiocarbon dating, aDNA, or stable isotopes. Most of the ZooMS samples measured between 1–2cm in length. A bone chip of ~20mg was removed from each bone fragment for ZooMS analysis using published ZooMS protocols (Brown et al. 2020; Buckley et al. 2009) (details on the applied ZooMS protocol can be found in the supplementary text). We identified the spectra using published ZooMS reference data (Buckley and Kansa 2011; Buckley et al. 2009; Janzen et al. 2021; Welker et al. 2016).

Despite being buried outside the cave for nearly 70 years, 85% of the 287 specimens preserved collagen for ZooMS identifications. Of the 287 samples, 202 were identified at least to the genus level, while 41 had more generic assignments (see Table 1). The ZooMS assemblage is much smaller than the plotted finds in previous studies; hence, no new species were identified. The ZooMS water-screened assemblage revealed a significantly higher number of hares, echoing the observations made by Boger et al. (2014). In addition, we discovered three new hominin fossils. All three fragments were small, weighing between 170mg and 407mg. Direct radiocarbon dating and aDNA analysis suggest they belonged to at least two individuals who lived in different periods, Magdalenian and Neolithic (Wang et al. authors' unpublished results).

We expected that the ZooMS data would illuminate the fragmentation patterns to some degree. To test this, we divided the 287 specimens into three weight groups: 95 specimens weighing between 70–250mg, 100 specimens weighing 250–600mg, and 92 specimens weighing >600mg (Figure 2). The lightest group includes most specimens that failed collagen extraction, indicating relatively poor COL1 preservation. Hares (*Lepus* sp.) mostly weigh less than 600mg, and bear (*Ursus*) are absent in the 70–250mg range, reflecting their respective body sizes. In contrast to our expectations, woolly rhinoceros (*Coelodonta antiquitatis*) are absent in the heaviest group. No qualitative correlation was observed between the counts and the specimen weights for dominant taxa such as horse, ungulate, and Elephantidae.

COMPARISON OF THE FAUNA DATASETS

The four zooarchaeological studies of Vogelherd presented here are distinct in their own ways. Lehmann's (1954) study focused only on the complete and identifiable portion of the 1931 fauna. Niven (2006) later re-examined the entire 1931 faunal collection using standard zooarchaeological approaches. Following the excavation of Vogelherd backdirt, Boger and colleagues (2014) analyzed the piece plotted fossils from the backdirt and compared them with the *in situ* assemblage. Here, we tested 287 bone fragments from the water-screened backdirt using ZooMS.

NISP and MNI are the fundamental quantitative units in zooarchaeology. NISP represents the most straightforward observational measure of taxonomic abundance, while MNI values derive from quantitative units depending on element identifications. There has been a long-standing debate regarding the advantages and disadvantages of

TABLE 2. COMPARISON OF HORSE (*Equus*) AND REINDER (*Rangifer tarandus*).*

	NISP				MNI		
	AH II-IX (Lehmann 1954)	AH II-IX (Niven 2006)	HL/KS (Boger et al. 2014)	HL/KS ZooMS (this study)	AH II-IX (Lehmann 1954)	AH II-IX (Niven 2006)	HL/KS (Boger et al. 2014)
Equus sp.	586	1825	229	45	52	61	13
Rangifer tarandus	115	1679	123	28	18	35	6
Equus / Rangifer sp. ratio	5.10	1.09	1.86	1.61	2.89	1.74	2.17

*AH means *in situ* archaeological horizon and HL/KS refers to the backdirt. Data from Boger et al. 2014; Lehmann 1954; Niven 2006.

MNI over NISP in zooarchaeological studies (Brothwell and Chaplin 1972; Domínguez-Rodrigo 2012; Grayson 1979; Lyman 2018; Marshall and Pilgram 1993; Morin et al. 2017; Uerpmann 1973). Most zooarchaeological studies list both NISP and MNI values in their taxonomic table (Niven 2006, 2007). However, there has been an emphasis shift from MNI to NISP values in the past decades (Lyman 2018). This trend is evident in the research of Vogelherd; Lehmann (1954) only listed MNI values in the taxonomic table, while Boger et al. (2014) listed NISP values only. This trend is compatible, and somehow beneficial to ZooMS studies since ZooMS identifications can only result in NISP counts.

Previous studies showed that although mammoth remains make up a large composition in the Vogelherd fauna, horses and reindeer are two primary prey taxa at Vogelherd. According to Niven (2007), Aurignacian people hunted both taxa seasonally, likely transporting complete carcasses back to the cave. The NISP and MNI values of horses and reindeer established in previous studies, and the ZooMS counts from this work, are used in a comparison (Table 2). As demonstrated previously (Lyman 2019), NISP values consistently correlate with MNI values across various studies—horses outnumber reindeer in both NISP and MNI (horse/reindeer ratios >1). Notably, while the data from Lehmann and Niven represent independent analyses of the same collection obtained from the 1931 excavation, their ratios display the largest difference (5.10 vs. 1.09). Conversely, assemblages derived from backdirt plotted and water-screened exhibit closer horse/reindeer ratios (1.86 vs. 1.61) (Boger et al. 2014; Lehmann 1954; Niven 2006).

In traditional zooarchaeological studies, body size classification, based on bone cortical thickness and fragment size (Brain 1983), is frequently used to group morphologically unidentifiable bone fragments. In the case of Vogelherd, both Niven (2006) and Boger et al. (2014) assigned unidentifiable specimens into body size groups and provided the tallies in their publications. Body size 5 includes mammoth and rhinoceros; body size 4, horse/bear/red deer; body size 3 includes reindeer/roe deer; and, body size 2 includes fox/hare (Boger et al. 2014; Niven, 2006). In Figure 3, we also convert the ZooMS result to body size groups for comparison. Based on more specific ZooMS assignments, the body size classification of the ZooMS assemblage is accurate and unbiased. We find that the body-size-based groups of Niven (2006) and Boger et al. (2014) (the outer and middle rings) show similar abundance patterns, especially for body sizes 3 and 4, suggesting that middle-sized animals dominated the non-diagnostic specimens. However, the ZooMS assemblage, which stems from highly fragmented specimens, shows a different pattern. Large-sized animals (body size 5) dominate the ZooMS assemblage and small-sized animals (body size 2) are also more abundant when compared to the other two datasets. The two body size patterns may be the result of actual differences between plotted and water-screened assemblages, or, as shown in other ZooMS-based studies, body size classes based on cortical thickness do not accurately reflect the overall species composition at a site (Sinet-Mathiot et al. 2019; 2023; Torres-Iglesias et al. 2024).

To better understand these differences, we convert the ZooMS dataset in taxonomic categories and compare it with the identifiable sub-assemblages from the previous three studies (Boger et al. 2014; Lehmann 1954; Niven 2006) (Figure 4). The broad categories, rather than precise taxa, are used in the comparison. Humans, hedgehogs (*Erinaceus europaeus*), and birds are absent in more than one dataset, and thus are excluded in this new comparison. Furthermore, NISP values from the three richest contexts, AH III, AH IV-V and AH VII, corresponding to the Magadalenian, Aurignacian and Middle Paleolithic, respectively, are extracted from Niven's dataset and shown in the bar chart of Figure 4a.

The assemblages of Niven (2006), Boger et al. (2014), and ZooMS have more complete categories, thus, we conducted a Chi-square test of independence between these three. The results show a significant difference in their

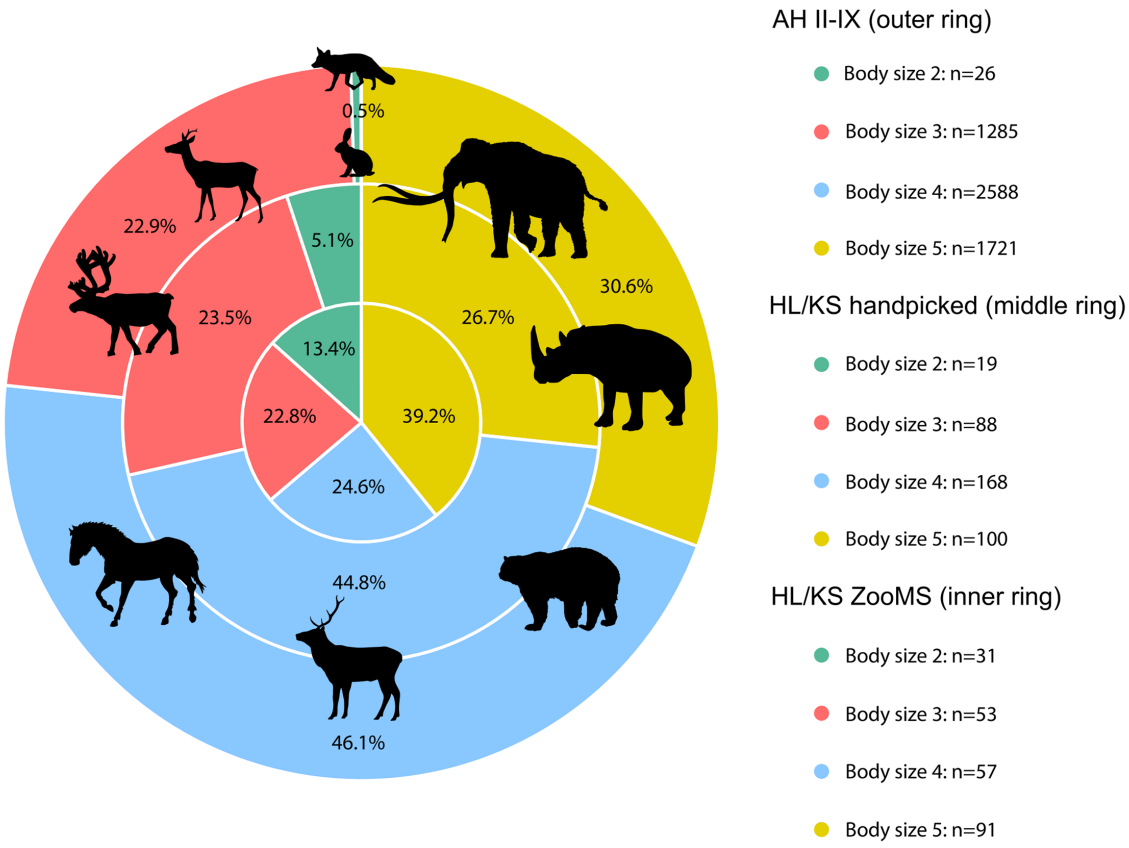


Figure 3. Body-size based groups and compositions of the morphologically unidentifiable bone remains at Vogelherd. The outer ring, AH II-IX, represents in situ data from Niven (2006); the middle ring shows data from the backdirt handpicked bones (Boger et al. 2014); the inner ring shows data from the ZooMS analysis of backdirt water-screened bones.

taxonomic profiles ($\chi 2 = 1924.6$, $df=14$, $p<0.01$), suggesting that the proportion of each category is not constant across assemblages, as shown in Figure 4a. Referring to the horse/reindeer ratio in these three assemblages, we would have expected the ZooMS taxonomic abundance to be closer to that of Boger et al.'s (2014) assemblage, as they are both based on materials from the backdirt. However, this is not the case. The chi-square standardized residuals (Figure 4b) of categories in Boger et al. (2014) are the highest and contribute the most to the difference. The ZooMS assemblage has the smallest residual contribution, which means it falls between the assemblages of Niven (2006) and Boger et al. (2014) and is more towards Niven's (2006).

Ivory and teeth fragments, both contributing significantly to the NISP values of mammoth and horse (67% and 52% in Niven's [2006] assemblage), are not part of the ZooMS dataset since only bones were studied. Hence, the similarity of the ZooMS-based assemblage with Niven's could be indeed higher. A distinct difference between the ZooMS and Niven's datasets is the identification of more hares and carnivores (except for bears) (see Figure 4b) in the ZooMS assemblage. We know that due to the speed of the operations and the focus on larger mammals and human remains, there were clear collecting biases against small animals in the 1931 excavation. Figure 2 has confirmed that hare bones are highly fragmented, mostly found in the 70–250mg weight group. Boger et al. (2014) also noted the large number of small game remains from the backdirt plotted assemblage, and they recorded clear human modifications on hare remains. Given the high fragmentation of hare remains in the water-screened assemblage, we suggest that small game, such as hares, may have played a more significant role in the subsistence strategies of the Aurignacian inhabitants of the site.

We should not expect an identical taxonomic abundance in ZooMS and morphologically identified assemblages. While coming from the same deposit, the ZooMS bones were small in size and morphologically undiagnostic. On the one hand, we can assume that counts achieved by ZooMS positively correlate with the body sizes of taxa (Brown et al. 2021). On the other hand, we may also suggest that intensive fragmentation is less likely to occur on larger mammal skeletons (Cannon 2013). Nevertheless, the abundance of a species in a zooarchaeological collection can never be assessed reliably in absolute terms (Lyman 2018).

Finally, we find Lehmann and Niven's abundance patterns unrelated, even after removing the mammoths from Niven's assemblage (Supplementary Material Figure S2).

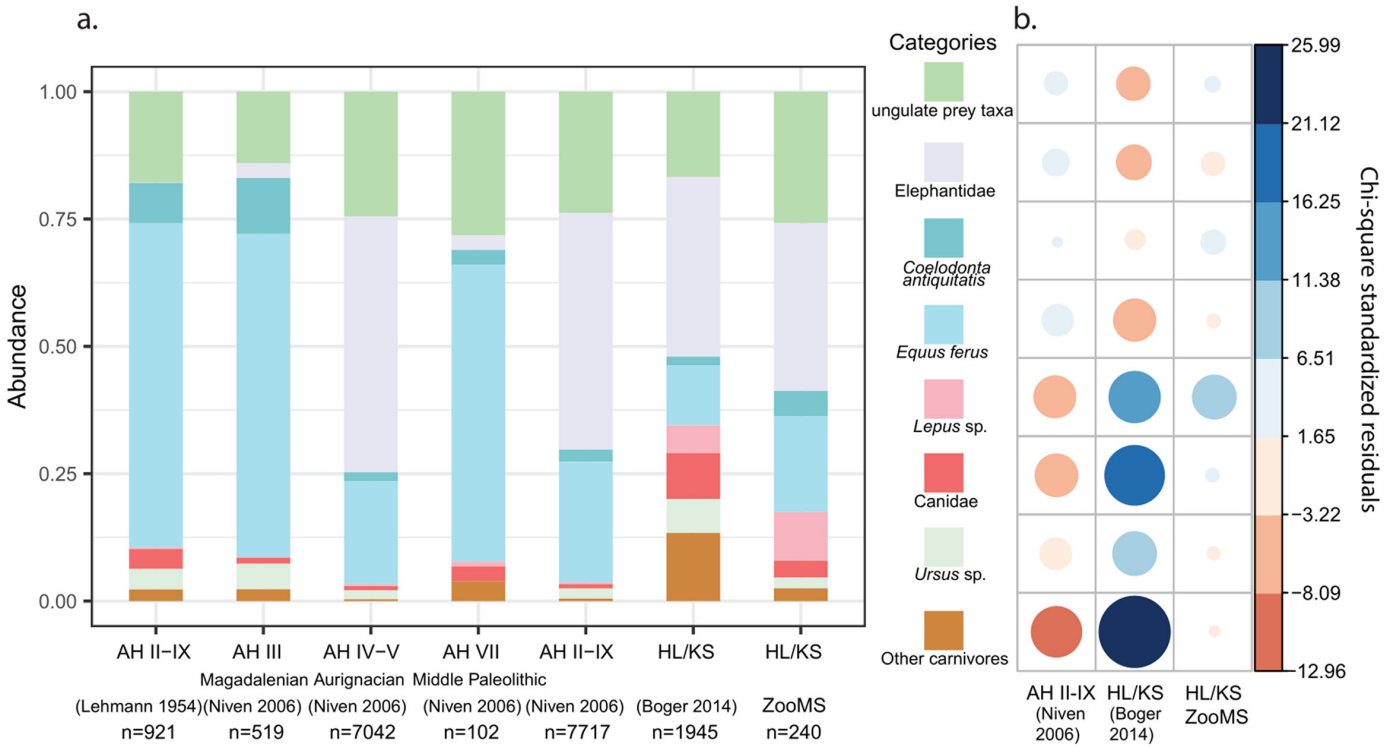


Figure 4. a) Abundance of identifiable bones by taxonomic categories, using NISP values. B) Visualization of chi-square standardized residuals of each category (data from Boger et al. 2014; Lehmann 1954; Niven 2006).

Although both assemblages are from the 1931 collection, Lehmann only examined a portion of it, while Niven examined the total of available fauna. This is a warning sign for the potential bias in partial sampling of an assemblage. Moreover, the work of Boger and colleagues stands out from the two previous analyses by a much higher identification rate (nearly 84%) of its assemblage.

DISCUSSION

ZooMS offers new opportunities for the taxonomic identification of traditionally undiagnostic bones. In recent years, diverse approaches to integrating ZooMS and zooarchaeological datasets have appeared in the literature (Brown et al. 2021; Ruebens et al. 2022; 2023; Silvestrini et al. 2022; Sinet-Mathiot et al. 2019; 2023). These case studies are tailored to the characteristics of the individual assemblages analyzed. However, as ZooMS data accumulates, methods for quantitative integration of assemblages analyzed using traditional and biomolecular approaches are becoming more pressing. At Vogelherd, four zooarchaeological studies were conducted using both morphological and molecular (ZooMS) approaches. Hence, the site can serve as a case study for exploring ways of integrating such datasets. Any integration attempt should rely on the understanding of both approaches, their benefits and limitations. In Table 3, we list a series of points regarding zooarchaeological analyses, mostly concerning Paleolithic sites, and then try to assess them from both ZooMS and traditional zooarchaeology perspectives.

Initial assessment. To evaluate the feasibility of ZooMS on a certain faunal collection, a small-scale test of collagen preservation is usually a priority. If the preservation allows, the analysis can be scaled up.

Whether in ZooMS or morphological approaches, researchers are concerned with the property of the studied assemblage. Usually, the ZooMS assemblage is the unidentifiable portion eliminated from the zooarchaeological morphological identification. Thus, a combination of the two approaches should bring a nearly complete understanding of taxonomic abundance. Sometimes, the situation may be more complex due to the intentional or unintentional preselection of samples. For example, in the Vogelherd ZooMS assemblage, ivory, teeth, burnt bones, and tiny bones <70mg are excluded from being analyzed. During the subsequent integration of datasets, we should consider the properties of distinct assemblages.

Preparation. Before examining a collection, compiling a list of potential taxa according to regional faunal studies is necessary for a zooarchaeologist. ZooMS researchers do the same listing, but mostly on a continental/biogeographical scale. Both ZooMS and morphological approaches rely on reference databases or collections; ultimately, our identifications are just as good as our reference collections are, and we can only determine a taxon if this exists in our comparative collection or reference database. For ZooMS, reference of common species is widely accessible and has less intra-species variation (at least for mammals) (Richter et al. 2022). On the other hand, zooarchaeological reference

TABLE 3. METHODOLOGICAL COMPARISON OF ZooMS AND TRADITIONAL ZOOARCHAEOLOGICAL APPROACHES.

	Questions (mainly related to Paleolithic sites)	ZooMS	Zooarchaeology (morphological approach)
Initial assessment	pilot study	small-scale testing to assess collagen preservation	not necessary
	specimen selection	>70mg (to allow for subsequent analysis if an interesting species is identified). Avoid sampling bone tools or worked bones with destructive protocols	often piece plotted finds, in most cases >3cm (except microfauna)
	sampling bias	preselection of small fragments or non-diagnostic fragments	entire or partial assemblage analyzed
Preparation	reference database	published or in-lab peptide markers for species identification	reference skeletal collections and publications
	possible species list	continental fauna list (more tolerance to unexpected species)	regional fauna list and reference skeletal collections
Identification	basis of ID	collagen peptide mass fingerprinting	characteristic morphological elements
	difficult samples	clustered by collagen proximity on amino acid sequence	clustered by morphological proximity
	ID taxonomic resolution	various among different families, generally family-genus level.	various among different families, generally species level or better
	ID success rate	depending on collagen preservation	depending on fragmentation level
	analyst's bias	minimal	experience-dependent
	processing time	high-throughput, hundreds of samples per week	experience-dependent
	open data and format	published MALDI-TOF spectra	published description, photos, 3D scan, and morphometric data
	quantification	direct ZooMS counts (NISP-like)	Quantification units: NISP, MNE, MAU, MNI etc.

collections offer an unparalleled diversity of (sub-)species accumulated by generations of scholars (Driver et al. 2011).

Identification. ZooMS outperforms traditional zooarchaeology in identifying bone fragments lacking diagnostic features, and the generally non-targeted reference database gives ZooMS an advantage in finding unexpected taxa. Regarding the accuracy of identification, the resolution of ZooMS is generally lower than that of morphological studies. However, with well-preserved COL1, ZooMS is able to differentiate morphologically similar species, e.g., sheep/goat, bison/buffalo, and donkey/horse (Buckley et al. 2010; Coutu et al. 2021; Jeanjean et al. 2023; Paladugu et al. 2023).

TABLE 3. METHODOLOGICAL COMPARISON OF ZooMS AND TRADITIONAL ZOOARCHAEOLOGICAL APPROACHES (continued).

	Questions (mainly related to Paleolithic sites)	ZooMS	Zooarchaeology (morphological approach)
Data interpretation	fragmentation level	correlation of taxa and specimen weights/lengths	bone circumference types and element survival pattern, or body size classification for non-IDed.
	accumulator of deposit	%carnivore or fragmentation patterns used as proxies	direct observations of gnawing/digestion marks, element survival pattern (e.g., %MAU), %carnivore, taxonomic diversity, age profile
	preservation	ID rate linked to collagen preservation and deamidation level an additional proxy for individual bone preservation	weathering state, color, root etching, density-mediated attrition
	sex & mortality profiles	rare	morphometric measurements
	environment adaption or phylogenetic evolution	difficult to detect from COL1 amino acids mutations	morphometric data (large dataset required)
	anthropogenic modifications	difficult to observe if bones are too small (<2cm)	cut marks, burning, breakage patterns
	spatial distribution of fauna remains in deposit	when no exact coordinates are available, spatial distribution can be based on larger units (squares, layers)	If piece-plotted, bone will have exact spatial coordinates. When no exact coordinates are available, spatial distribution will be based on larger units (squares, layers).

It is difficult for the morphological approach unless using a large biometrics dataset for local species (Hanot and Bochaton 2018; Horsburgh 2018; Scott and Plug 2016).

ZooMS identification is not equally effective for all mammals. In ZooMS identification, a widely accepted reference system comprises 12 peptide markers—some are more detectable, others are more diagnostic. That is why ZooMS markers have different levels of effectiveness for identification. These discrepancies in identifiability are similar to those in morphological identification (Driver et al. 2011; Lyman 2015b; Wong et al. 2017).

Both ZooMS and morphological approaches may reach assignments broader than the genus. It is due to the inability to locate or differentiate specific morphological features (Lau and Kansa 2018; Lyman 2015b), or the lack of diagnostic peptide markers in ZooMS spectra. Hence, the term "taxon" is used as and is not restricted to genus/species (Lyman 1994).

An advantage of ZooMS over comparative zooarchaeology is the standardized method for data analyses. With ZooMS, bones can be identified "each on its own merits" without many assumptions imposed by the analysts. On the contrary, the quality of a morphological analysis depends more on the analyst's experience. Zooarchaeologists have noticed the inter-analyst variability for decades (Domínguez-Rodrigo 2012; Gobalet 2001; Lau and Kansa

2018). While initiatives like the blind test and quality control in the morphological analysis were proposed (Driver et al. 2011; Wolverton 2013), zooarchaeologists rarely undertake these in practice.

Although not entirely absolute, a ZooMS identification depends much less on the analyst's experience. Most ZooMS analysts process spectra following a standardized workflow (Brown 2021), if not automatic programming (Gu and Buckley 2018). In addition, it is easy to clarify specific criteria used to identify one taxon over the other. Nevertheless, a recent double-blind experiment has verified that ZooMS and morphological identification yield statistically indistinguishable taxonomic profiles for the same archaeological faunal assemblage (Morin et al. 2023). Ultimately, ZooMS can contribute to quality control of morphological analysis as long as collagen preservation allows.

Data interpretation. Following taxonomic assignments, the resulting dataset forms the basis for interpreting specific questions linked to the regional ecology, paleoclimate and human subsistence patterns. Morphological identification relies on the identity of the skeletal elements. Thus, the taxonomic results naturally include elemental composition and morphometric information. In contrast, ZooMS identification depends on the presence of diagnostic collagen peptides, which can reflect the biomolecular preservation of samples (Wang et al. 2021). Element survival pattern is essential in reconstructing human subsistence and is statistically quantitative. The lack of elemental information disadvantages faunal assemblages sieved from the sediment, like the Vogelherd ZooMS assemblage. However, in a few recent integration attempts, ZooMS assemblages have incorporated specimens representing specific anatomical elements or having traces of modification, while lacking precise taxonomic identifications (Pothier-Bouchard et al. 2020; Ruebens et al. 2023; Silvestrini et al. 2022; Torres-Iglesias et al. 2024).

ZooMS greatly augments the number of specimens that can be assigned to taxa. Meanwhile, the weight (or length) of each single specimen could be routinely measured before ZooMS lab work due to the generally smaller size of specimens in ZooMS assemblage. Combining weight (or length) data and expanded taxa identifications, ZooMS can provide powerful insights into bone fragmentation, preservation, and spatial distribution. Integrated with zooarchaeological results, ZooMS allows for a more comprehensive archaeological interpretation.

New proteomic approaches like SPIN with higher taxonomic resolution offer hope to replace ZooMS, providing higher quality and more in-depth results to zooarchaeology (Rüther et al. 2022). Ultimately, the integration of proteomic approaches, environmental sediment DNA (e.g., Zavala et al. 2021) and traditional zooarchaeology on species abundance will transform how we understand the faunal record of an archaeological or paleontological site.

CONCLUSIONS

In the new era of ample scientific research in archaeology, a central goal is the combination of all analytical tools to represent and understand the past better. As often quoted, scientific interpretations are only probable reconstructions of reality, and usually simply empirical approximations, rather than the absolute reality of what once was (Bunge 1998).

In this study, we use the history of research on the fauna of Vogelherd as a case study for understanding the change of research focus in archaeological fauna remains. The application of ZooMS represents the most recent analytical tool used at Vogelherd, and this method would greatly augment the number of specimens that can now be identified to the taxa. Here, we compiled the data from all previous and current studies and compared the datasets, in an attempt to provide guidelines for integrating and interpreting zooarchaeological data gained from morphological and ZooMS approaches.

While our interpretations remain an "empirical approximation" of past ecology and human behavior, we hope that by combining and integrating such datasets within a concise framework, researchers in the coming years will significantly expand our understanding of the archaeological record in ways that remained out of reach before the advent of ZooMS.

DATA AVAILABILITY

We have uploaded all ZooMS spectra and results files to Mendeley data. ("Vogelherd ZooMS data," doi: 10.17632/9jp4jdzy7k.1)

ACKNOWLEDGEMENTS

This work received funding from the ERC under the European Union's Horizon 2020 Research and Innovation Programme grant agreement no. 715069 (FINDER Project) to K.D. N.W. received a doctoral scholarship from the FINDER Project. The excavations at Vogelherd have been funded by the Deutsche Forschungsgemeinschaft SFB ResourcenKulturen, the Heidelberg Akademie der Wissenschaften long-term project the Role of Culture in the Early Expansion of Modern Humans, The Ministry of Science of the State of Baden-Württemberg, the University of Tübingen, the Verein für Eiszeitkunst im Lonetal, the Hans Voith Foundation, the city of Neiderstotzingen, Kreis Heidenheim and Heidelberger Materials. Among the scores of excavators and people processing faunal materials from Vogelherd, we are particularly indebted to Laura Niven, Britt Starkovich, Susanne Münzel, Angel Blanco-Lapez, Maria Malina, Mohsen Zeidi, Alexander Janas, and Saman Hamzavi. We also thank Michelle O'Reilly for assistance with the figure design.

REFERENCES

Andersen, K.K., Svensson, A., Johnsen, S.J., 2006. The Greenland ice core chronology 2005, 15–42 ka. Part 1:

constructing the time scale. Quat. Sci. Rev. 25, 23–24.
Banning, E.B., 2020. Archaeological animal remains, in: Banning, E.B. (Ed.), The Archaeologist's Laboratory: The Analysis of Archaeological Evidence. Springer International Publishing, Cham, pp. 241–266.
Behrensmeyer, A.K., Gordon, K.D., Yanagi, G.T., 1986. Trampling as a cause of bone surface damage and pseudo-cutmarks. Nature 319, 768–771.
Binford, L.R., 1981. Bones: Ancient Men and Modern Myths. Academic Press, New York.
Boger, U., Starkovich, B.M., Conard, N.J., 2014. New insights gained from the faunal material recovered during the latest excavations at Vogelherd Cave. Mitteilungen der Gesellschaft für Urgeschichte 23, 57–81.
Brain, C.K., 1983. The Hunters Or the Hunted?: An Introduction to African Cave Taphonomy. University of Chicago Press, Chicago.
Brown, S., 2021. Identifying ZooMS Spectra (mammals) using mMass. https://dx.doi.org/10.17504/protocols.io.bzscp6aw
Brown, S., Hebestreit, S., Wang, N., Boivin, N., Douka, K., 2020. Zooarchaeology by Mass Spectrometry (ZooMS) for bone material-AmBiC protocol. https://dx.doi.org/10.17504/protocols.io.bffdjji6
Brown, S., Wang, N., Oertle, A., Kozlikin, M.B., Shunkov, M.V., Derevianko, A.P., Comeskey, D., Jope-Street, B., Harvey, V.L., Chowdhury, M.P., Buckley, M., Higham, T., Douka, K., 2021. Zooarchaeology through the lens of collagen fingerprinting at Denisova Cave. Sci. Rep. 11, 15457.
Brumley, J.H., 1973. Quantitative methods in the analysis of butchered faunal remains: a suggested approach. Archaeol. Montana 14(1), 1–40.
Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. 23, 3843–3854.
Buckley, M., Harvey, V.L., Chamberlain, A.T., 2017. Species identification and decay assessment of Late Pleistocene fragmentary vertebrate remains from Pin Hole Cave (Creswell Crags, UK) using collagen fingerprinting. Boreas 46, 402–411.
Buckley, M., Kansa, S.W., 2011. Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. Archaeol. Anthropol. Sci. 3, 271–280.
Buckley, M., Whitcher Kansa, S., Howard, S., Campbell, S., Thomas-Oates, J., Collins, M., 2010. Distinguishing between archaeological sheep and goat bones using a single collagen peptide. J. Archaeol. Sci. 37, 13–20.
Bunge, M., 1998. Philosophy of Science: From Explanation to Justification. Transaction Publishers, New Brunswick.
Cannon, M.D., 2013. NISP, Bone fragmentation, and the measurement of taxonomic abundance. J. Archaeol. Method Theory 20, 397–419.
Chaplin, R.E., 1971. The Study of Animal Bones from Archaeological Sites. Seminar Press, London.
Conard, N.J., 2009. Jünger als gedacht! Zur Neudatierung der Menschenreste vom Vogelherd. In: Archäologisches Landesmuseum Baden-Württemberg/Abteilung Ältere Urgeschichte und Quartärökologie der Eberhard Karls Universität Tübingen (Eds.), Begleitband zur Großen Landesausstellung Eiszeit – Kunst und Kultur im Kunstgebäude Stuttgart. Jan Thorbecke Verlag, Ostfildern.
Conard, N.J., Bolus, M., 2003. Radiocarbon dating the appearance of modern humans and timing of cultural innovations in Europe: new results and new challenges. J. Hum. Evol. 44, 331–371.
Conard, N.J., Grootes, P.M., Smith, F.H., 2004. Unexpectedly recent dates for human remains from Vogelherd. Nature 430, 198–201.
Conard, N.J., Janas, A., Zeidi, M., 2015a. Neues aus dem Lonetal: Ergebnisse von Ausgrabungen an der Fetzershaldenhöhle und dem Vogelherd. Archäo. Ausgrabungen Baden-Würrtemberg 2014, 59–64.
Conard, N.J., Zeidi, M., Bega, J., 2012. Die letzte Kampagne der Nachgrabungen am Vogelherd. Archäologische Ausgrabungen in Baden-Württemberg. Archäol. Ausgrabungen Baden-Württemberg 2012, 84–88.
Conard, N.J., Zeidi, M., Janas, A., 2015b. Abschließender Bericht über die Nachgrabung am Vogelherd und die Sondage in der Wolftalhöhle. Archäol. Ausgrabungen Baden-Württemberg 2014, 66–72.
Cornwall, I.W., 1956. Bones for the Archaeologist. Phoenix House, London.
Coutu, A.N., Taurozzi, A.J., Mackie, M., Jensen, T.Z.T., Collins, M.J., Sealy, J., 2021. Palaeoproteomics confirm earliest domesticated sheep in southern Africa ca. 2000 BP. Sci. Rep. 11, 6631.
Cuijpers, A.G.F.M., 2006. Histological identification of bone fragments in archaeology: telling humans apart from horses and cattle. Intern. J. Osteoarchaeol. 16, 465–480.
Domínguez-Rodrigo, M., 2012. Critical review of the MNI (minimum number of individuals) as a zooarchaeological unit of quantification. Archaeol. Anthropol. Sci. 4, 47–59.
Driver, J.C., Bovy, K., Butler, V.L., Lupo, K.D., Lyman, L.R., Otaola, C., 2011. Identification, classification and zooarchaeology. Ethnobiol. Lett. 2, 19–39.
Eaton, G.F., 1898. The prehistoric fauna of Block Island, as indicated by its ancient shell-heaps. Am. J. Sci. 6, 137–161.
Forchhammer, G., 1852. Undersögelser i geologisk -antiqvarisk Retning. Biance Luno, Copenhagen.
Geiling, J.M., Marín-Arroyo, A.B., Straus, L.G., González Morales, M.R., 2018. Deciphering archaeological palimpsests with bone micro-fragments from the Lower Magdalenian of El Mirón cave (Cantabria, Spain). Hist. Biol. 30, 730–742.
Giovas, C.M., LeFebvre, M.J., 2017. Zooarchaeology in Practice: Case Studies in Methodology and Interpretation in Archaeofaunal Analysis, 1st ed. Springer International Publishing, Cham, Switzerland.

Gobalet, K.W., 2001. A critique of faunal analysis; inconsistency among experts in blind tests. J. Archaeol. Sci. 28, 377–386.

Grayson, D.K., 1979. On the quantification of vertebrate archaeofaunas. Advances Archaeol. Method Theory 2, 199–237.

Grealy, A.C., McDowell, M.C., Scofield, P., Murray, D.C., Fusco, D.A., Haile, J., Prideaux, G.J., Bunce, M., 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages. Quat. Sci. Rev. 128, 37–47.

Grigson, C., 2016. Towards a blueprint for animal bone reports in archaeology. Research problems in zooarchaeology. In: Brothwell, D.R., Kenneth D.T., Clutton-Brock, J. (Eds.), Research Problems in Zooarchaeology. Routledge. New York, pp. 121–128. https://doi.org/10.4324/9781315421094-16/towards-blueprint-animal-bone-reports-archaeology-caroline-grigson

Gu, M., Buckley, M., 2018. Semi-supervised machine learning for automated species identification by collagen peptide mass fingerprinting. BMC Bioinformatics 19, 241.

Hahn, J., 1988. Die Geißenklösterle-Höhle im Achtal bei Blaubeuren I: Fundhorizontbildung und Besiedlung im Mittelpaläolithikum und im Aurignacien. Theiss, Stuttgart.

Hahn, J., 1977. Aurignacien, das ältere Jungpaläolithikum in Mittel-und Osteuropa. Böhlau, Cologne.

Hanot, P., Bochaton, C., 2018. New osteological criteria for the identification of domestic horses, donkeys and their hybrids in archaeological contexts. J. Archaeol. Sci. 94, 12–20.

Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C.B., Conard, N.J., 2012. Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: the radiocarbon chronology of Geißenklösterle. J. Hum. Evol. 62, 664–676.

Horsburgh, A.K., 2018. A reply to Plug 2017: science requires self-correction. Azania: Archaeol. Res. Afr. 53, 114–118.

Horsburgh, K.A., 2008. Wild or domesticated? An ancient DNA approach to canid species identification in South Africa's Western Cape Province. J. Archaeol. Sci. 35, 1474–1480.

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

Jeanjean, M., McGrath, K., Valenzuela-Lamas, S., Nieto-Espinet, A., Schafberg, R., Parés-Casanova, P.M., Jiménez-Manchón, S., Guintard, C., Tekkouk, F., Ridouh, R., Mureau, C., Evin, A., 2023. ZooMS confirms geometric morphometrics species identification of ancient sheep and goat. R Soc Open Sci 10, 230672.

Johnson, E., 1985. Current developments in bone technology. Advances Archaeol. Method Theory 8, 157–235.

Kitagawa, K., 2014. Exploring Hominins and Animals in the Swabian Jura: Study of the Paleolithic Fauna from Hohlenstein-Stadel. Ph.D. Dissertation. Eberhard Karls Universität Tübingen.

Lau, H., Kansa, S.W., 2018. Zooarchaeology in the era of big data: Contending with interanalyst variation and best practices for contextualizing data for informed reuse. J. Archaeol. Sci. 95, 33–39.

Lawrence, B., 1951. … Awatovi Site Part II. Post-Cranial Skeletal Characters of Deer, Pronghorn, and Sheep-Goat with Notes on *Bos* and *Bison*. Peabody Museum Archaeol. Ethnol. 35(3), 9–43.

Lehmann, U., 1954. Die Fauna des "Vogelherds" bei Stetten ob Lontal (Württemberg). Neues Jahrb. Geol. Paläontol 99, 33–146.

Librado, P., Khan, N., Fages, A., Kusliy, M.A., Suchan, T., Tonasso-Calvière, L., Schiavinato, S., Alioglu, D., Fromentier, A., Perdereau, A., Aury, J.-M., Gaunitz, C., Chauvey, L., Seguin-Orlando, A., Der Sarkissian, C., Southon, J., Shapiro, B., Tishkin, A.A., Kovalev, A.A., Alquraishi, S., Alfarhan, A.H., Al-Rasheid, K.A.S., Seregély, T., Klassen, L., Iversen, R., Bignon-Lau, O., Bodu, P., Olive, M., Castel, J.-C., Boudadi-Maligne, M., Alvarez, N., Germonpré, M., Moskal-Del Hoyo, M., Wilczyński, J., Pospuła, S., Lasota-Kuś, A., Tunia, K., Nowak, M., Rannamäe, E., Saarma, U., Boeskorov, G., Lõugas, L., Kyselý, R., Peške, L., Bălăşescu, A., Dumitraşcu, V., Dobrescu, R., Gerber, D., Kiss, V., Szécsényi-Nagy, A., Mende, B.G., Gallina, Z., Somogyi, K., Kulcsár, G., Gál, E., Bendrey, R., Allentoft, M.E., Sirbu, G., Dergachev, V., Shephard, H., Tomadini, N., Grouard, S., Kasparov, A., Basilyan, A.E., Anisimov, M.A., Nikolskiy, P.A., Pavlova, E.Y., Pitulko, V., Brem, G., Wallner, B., Schwall, C., Keller, M., Kitagawa, K., Bessudnov, A.N., Bessudnov, A., Taylor, W., Magail, J., Gantulga, J.-O., Bayarsaikhan, J., Erdenebaatar, D., Tabaldiev, K., Mijiddorj, E., Boldgiv, B., Tsagaan, T., Pruvost, M., Olsen, S., Makarewicz, C.A., Valenzuela Lamas, S., Albizuri Canadell, S., Nieto Espinet, A., Iborra, M.P., Lira Garrido, J., Rodríguez González, E., Celestino, S., Olària, C., Arsuaga, J.L., Kotova, N., Pryor, A., Crabtree, P., Zhumatayev, R., Toleubaev, A., Morgunova, N.L., Kuznetsova, T., Lordkipanize, D., Marzullo, M., Prato, O., Bagnasco Gianni, G., Tecchiati, U., Clavel, B., Lepetz, S., Davoudi, H., Mashkour, M., Berezina, N.Y., Stockhammer, P.W., Krause, J., Haak, W., Morales-Muñiz, A., Benecke, N., Hofreiter, M., Ludwig, A., Graphodatsky, A.S., Peters, J., Kiryushin, K.Y., Iderkhangai, T.-O., Bokovenko, N.A., Vasiliev, S.K., Seregin, N.N., Chugunov, K.V., Plasteeva, N.A., Baryshnikov, G.F., Petrova, E., Sablin, M., Ananyevskaya, E., Logvin, A., Shevnina, I., Logvin, V., Kalieva, S., Loman, V., Kukushkin, I., Merz, I., Merz, V., Sakenov, S., Varfolomeyev, V., Usmanova, E., Zaibert, V., Arbuckle, B., Belinskiy, A.B., Kalmykov, A., Reinhold, S., Hansen, S., Yudin, A.I., Vybornov, A.A., Epimakhov, A., Berezina, N.S., Roslyakova, N., Kosintsev, P.A.,

Kuznetsov, P.F., Anthony, D., Kroonen, G.J., Kristiansen, K., Wincker, P., Outram, A., Orlando, L., 2021. The origins and spread of domestic horses from the Western Eurasian steppes. Nature 598, 634–640.

Lyman, L.R., 1994. Quantitative units and terminology in zooarchaeology. Am. Antiq. 59, 36–71.

Lyman, L.R., 2015a. The history of "laundry lists" in North American zooarchaeology. J. Anthropol. Archaeol. 39, 42–50.

Lyman, L.R., 2015b. On the variable relationship between NISP and NTAXA in bird remains and in mammal remains. J. Archaeol. Sci. 53, 291–296.

Lyman, L.R., 2018. Observations on the history of zooarchaeological quantitative units: Why NISP, then MNI, then NISP again? J. Archaeol. Sci. Rep. 18, 43–50.

Lyman, L.R., 2019. A critical review of four efforts to resurrect MNI in zooarchaeology. J. Archaeol. Method Theory 26, 52–87.

Marean, C.W., 1998. A critique of the evidence for scavenging by Neanderthals and early modern humans: new data from Kobeh Cave (Zagros Mountains, Iran) and Die Kelders Cave 1 layer 10 (South Africa). J. Hum. Evol. 35, 111–136.

Marean, C.W., Rodrigo, M.D., Pickering, T.R., 2004. Skeletal element equifinality in zooarchaeology begins with method: the evolution and status of the "Shaft Critique." J. Taphonomy 2, 69–98.

Marshall, F., Pilgram, T., 1993. NISP vs. MNI in quantification of body-part representation. Am. Antiq. 58, 261–269.

Martisius, N.L., Spasov, R., Smith, G.M., Endarova, E., Sinet-Mathiot, V., Welker, F., Aldeias, V., Horta, P., Marreiros, J., Rezek, Z., McPherron, S.P., Sirakov, N., Sirakova, S., Tsanova, T., Hublin, J.-J., 2022. Initial Upper Paleolithic bone technology and personal ornaments at Bacho Kiro Cave (Bulgaria). J. Hum. Evol. 167, 103198.

Mills, W.C., 1904. Exploration of the Gartner mound and village site. Ohio State Archaeol. Hist. Quarterly 13, 128–189.

Morin, E., Oldfield, E.-M., Bakovic, M., Bordes, J.-G., Castel, J.-C., Crevecoeur, I., Monnier, G., Tostevin, G., Buckley, M., 2023. A double-blind comparison of morphological and collagen fingerprinting (ZooMS) methods of skeletal identifications from Paleolithic contexts. Sci. Rep. 13, 18825. https://doi.org/10.21203/rs.3.rs-3083727/v1

Morin, E., Ready, E., Boileau, A., Beauval, C., Coumont, M.-P., 2017. Problems of identification and quantification in archaeozoological analysis, part I: insights from a blind test. J. Archaeol. Method Theory 24, 886–937.

Müller-Beck, H., 1957. Das obere Altpaläolithikum in Süddeutschland : ein Versuch zur ältesten Geschichte des Menschen. Habelt, Bonn.

Niven, L., 2001. The role of mammoths in Upper Palaeolithic economies of southern Germany. In: Cavarretta, G., Gioia, P., Mussi, M., Palombo, M. (Eds.), The World of Elephants: Proceedings of the 1st International Congress. CNR, Rome, pp. 323–327. https://sovraintendenzaroma.it/sites/default/files/storage/original/application/b0aa86ed32af7a62cfb0b69cc049e6e5.pdf (accessed 2.15.23).

Niven, L., 2003. Patterns of subsistence and settlement during the Aurignacian of the Swabian Jura, Germany. In: Zilhão, J., d'Errico, F. (Eds.), The Chronology of the Aurignacian and of the Transitional Complexes: Dating, Stratigraphies, Cultural Implications. Trabalhos de Arqueologia. Lisbon, pp. 199–211. https://www.patrimoniocultural.gov.pt/media/uploads/trabalhosdearqueologia/33/15.pdf (accessed 2.15.23).

Niven, L., 2006. The Palaeolithic Occupation of Vogelherd Cave : Implications for the Subsistence Behavior of Late Neanderthals and Early Modern Humans. Kerns, Tübingen.

Niven, L., 2007. From carcass to cave: large mammal exploitation during the Aurignacian at Vogelherd, Germany. J. Hum. Evol. 53, 362–382.

Olsen, S.J., 1971. Zooarchaeology: Animal Bones in Archaeology and Their Interpretation. Addison-Wesley Publishing Company, Boston.

Paladugu, R., Richter, K.K., Valente, M.J., Gabriel, S., Detry, C., Warinner, C., Dias, C.B., 2023. Your horse is a donkey! Identifying domesticated equids from Western Iberia using collagen fingerprinting. J. Archaeol. Sci. 149, 105696.

Pothier-Bouchard, G., Riel-Salvatore, J., Negrino, F., Buckley, M., 2020. Archaeozoological, taphonomic and ZooMS insights into the Protoaurignacian faunal record from Riparo Bombrini. Quat. Int. 551, 243–263.

Rasmussen, S.O., Bigler, M., Blockley, S.P., Blunier, T., Buchardt, S.L., Clausen, H.B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S.J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W.Z., Lowe, J.J., Pedro, J.B., Popp, T., Seierstad, I.K., Steffensen, J.P., Svensson, A.M., Vallelonga, P., Vinther, B.M., Walker, M.J.C., Wheatley, J.J., Winstrup, M., 2014. A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. Quat. Sci. Rev. 106, 14–28.

Reitz, E.J., Wing, E.S., 1999. Zooarchaeology. Cambridge University Press, New York.

Richter, K.K., Codlin, M.C., Seabrook, M., Warinner, C., 2022. A primer for ZooMS applications in archaeology. Proc. Natl. Acad. Sci. U.S.A. 119, e2109323119.

Riek, G., 1934. Die Eiszeitjägerstation am Vogelherd im Lonetal. Akademische Verlagsbuchhandlung Franz F. Heine, Tübingen.

Ruebens, K., Sinet-Mathiot, V., Talamo, S., Smith, G.M., Welker, F., Hublin, J.-J., McPherron, S.P., 2022. The late Middle Palaeolithic occupation of Abri du Maras (Layer 1, Neronian, Southeast France): integrating lithic analyses, ZooMS and radiocarbon dating to reconstruct Neanderthal hunting behaviour. J. Paleolit. Archaeol. 5, 4.

Ruebens, K., Smith, G.M., Fewlass, H., Sinet-Mathiot, V., Hublin, J.-J., Welker, F., 2023. Neanderthal subsistence, taphonomy and chronology at Salzgitter-Lebenstedt

(Germany): a multifaceted analysis of morphologically unidentifiable bone. J. Quat. Sci. 38(4), 471–487. https://doi.org/10.1002/jqs.3499

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

Schuerch, B., Wolf, S., Schmidt, P., Conard, N.J., 2021. Molluscs of the genus *Glycymeris* from Vogelherd cave near Niederstotzingen (Lonetal, southwestern Germany). MGfU 29. Mitt. Ges. Urgesch. 29, 53–79. https://doi.org/10.51315/mgfu.2020.29003

Schürch, B., Wong, L.G., Luzi, E., Conard, N.J., in prep. Evidence for an earlier Magdalenian presence in the Lone Valley of southwest Germany. Available at SSRN: https://ssrn.com/abstract=4712167

Scott, K., Plug, I., 2016. Osteomorphology and osteometry versus aDNA in taxonomic identification of fragmentary sheep and sheep/goat bones from archaeological deposits …. S. Afr. Humanities 28, 61–79.

Silvestrini, S., Lugli, F., Romandini, M., Real, C., Sommella, E., Salviati, E., Arrighi, S., Bortolini, E., Figus, C., Higgins, O.A., Marciani, G., Oxilia, G., Delpiano, D., Vazzana, A., Piperno, M., Crescenzi, C., Campiglia, P., Collina, C., Peresani, M., Spinapolice, E.E., Benazzi, S., 2022. Integrating ZooMS and zooarchaeology: new data from the Uluzzian levels of Uluzzo C Rock Shelter, Roccia San Sebastiano cave and Riparo del Broion. PLoS One 17, e0275614.

Sinet-Mathiot, V., Rendu, W., Steele, T.E., Spasov, R., Madelaine, S., Renou, S., Soulier, M.-C., Martisius, N.L., Aldeias, V., Endarova, E., Goldberg, P., McPherron, S.J.P., Rezek, Z., Sandgathe, D., Sirakov, N., Sirakova, S., Soressi, M., Tsanova, T., Turq, A., Hublin, J.-J., Welker, F., Smith, G.M., 2023. Identifying the unidentified fauna enhances insights into hominin subsistence strategies during the Middle to Upper Palaeolithic transition. Archaeol. Anthropol. Sci. 15, 139.

Sinet-Mathiot, V., Smith, G.M., 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). Sci. Rep. 9, 12350.

Torres-Iglesias, L., Marín-Arroyo, A.B., Welker, F., 2024. Using ZooMS to assess archaeozoological insights and unravel human subsistence behaviour at La Viña rock shelter (northern Iberia). J. Archaeol. Sci., 161, 105904.

Uerpmann, H.P., 1973. Animal bone finds and economic archaeology: a critical study of "osteo-archaeological" method. World Archaeol. 4, 307–322.

Wang, N., Brown, Samantha., Ditchfield, P., Hebestreit, S., Kozilikin, M., Luu, S., Wedage, O., Grimaldi, S., Chazan, M., Horwitz, H.L., Spriggs, M., Summerhayes, G., Shunkov, M., Richter, K.K., Douka, K., 2021. Testing the efficacy and comparability of ZooMS protocols on archaeological bone. J. Proteomics 233, 104078.

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, B.M., Stewart, J.R., Pääbo, S., Collins, M.J., Hublin, J.-J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proc. Natl. Acad. Sci. U.S.A. 113, 11162–11167.

Wolf, S., Conard, N.J., 2015. Personal ornaments of the Swabian Aurignacian. Palethnologie 7, 61–79. https://doi.org/10.4000/palethnologie.897

Wolverton, S., 2013. Data Quality in zooarchaeological faunal identification. J. Archaeol. Method Theory 20, 381–396.

Wong, G.L., Starkovich, B.M., Conard, N.J., 2017. Human subsistence and environment during the Magdalenian at Langmahdhalde: evidence from a new rock shelter in the Lone Valley, southwest Germany. Mitteilungen der Gesellschaft für Urgeschichte 26, 103–123.

Wright, D., Hughes, P., Skopal, N., Kmošek, M., Way, A., Sullivan, M., Lisá, L., Ricardi, P., Škrdla, P., Nejman, L., Gadd, P., Fišáková, M.N., Mlejnek, O., Králík, M., 2021. The archaeology of overburden: method within the madness at Švédův Stůl, Czech Republic. J. Archaeol. Sci. 132, 105429.

Special Issue: Integrating ZooMS and Zooarchaeology:
Methodological Challenges and Interpretive Potentials

Supplement: Integrating Morphological and ZooMS-Based Approaches to Zooarchaeology at Vogelherd Cave in Southwestern Germany

NAIHUI WANG
Max Planck Institute of Geoanthropology, 07745 Jena; and, Department of Early Prehistory and Quaternay Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen, GERMANY; nwang@gea.mpg.de

NICHOLAS J. CONARD
Department of Early Prehistory and Quaternay Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen; and, Senckenberg Centre for Human Evolution and Palaeoenvironment, Univesity of Tübingen, 72070 Tübingen, GERMANY; nicholas.conard@uni-tuebingen.de

KATERINA DOUKA
Max Planck Institute of Geoanthropology, 07745 Jena, GERMANY; Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, 1030 Vienna; and, Human Evolution and Archaeological Sciences (HEAS), University of Vienna, 1030 Vienna, AUSTRIA; katerina.douka@univie.ac.at

SUPPLEMENT

This supplement includes: supplementary material text, supplementary material figures S1–S2, and supplementary material references. ZooMS spectra files (in .mzml formats) and results are accessible on Mendeley data ("Vogelherd ZooMS data," doi: 10.17632/9jp4jdzy7k.1).

ZOOMS SAMPLE PREPARATION AND DATA ANALYSIS

We sampled 276 bones by removing a 20mg chip from each, and these were originally treated using the Ambic protocol. Then 114 samples that fail to yield ideal spectra, together with an additional 11 samples, were performed with acid-insoluble protocol as below.

AmBic protocol. Each bone sample was covered in 100µL 50 mM ammonium bicarbonate (NH_4HCO_3) at room temperature overnight, to clean and remove soluble contamination. The supernatant was discarded and an additional 100µL of 50 mM NH_4HCO_3 was added in. Following incubation at 65°C for 1 h, the bone chips were frozen at -20°C and the supernatant was digested with 0.4µg trypsin at 37°C for 18 h. After that, 1µL 5% trifluoroacetic acid (TFA) was added to end the digestion. The resulting supernatant was concentrated and desalted using C18 ZipTips, then washed with 200µL 0.1% TFA and eluted with 50µL 50% ACN/ 0.1% TFA (*v*/*v*).

Acid-insoluble protocol. The bone chips were demineralized in 500µL 0.5 M HCl for 24-48 h at 4°C until the bone chips became spongy and stopped reacting. The acid supernatant was removed. The chips were then rinsed 3 times using 0.5M NH_4HCO_3 until a neutral pH was reached. The samples were incubated at 65 °C for 1 h, in 100µL of 50 mM NH_4HCO_3. After the incubation, 50µL supernatant was digested and desalted as above.

All tryptic extracts were diluted 10 times with 50% ACN/0.1% TFA (*v*/*v*), and mixed with an equal volume of α -cyano-4-hydroxycinnamic acid solution (10mg/mL in 50% ACN/0.1% TFA (*v*/*v*)). 1.5µL of the mixture was spotted on a Bruker ground steel plate in triplicate. One blank was analyzed alongside every twenty-three samples as a negative control. Samples were measured using an Autoflex Speed LRF Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometer (Bruker).

Mass spectra files were processed using the mMass open software version 5.5.0 (Strohalm et al. 2010). ZooMS identification is done with an in-room tool. Previously published type I collagen peptide markers were used for the taxonomic identification of each sample (Buckley and Kansa 2011; Buckley et al. 2009; Janzen et al. 2021; Welker et al. 2016).

Statistical analysis was conducted in R (Team and Others 2013) and figures were produced using the package ggplot2 (Wickham 2016).

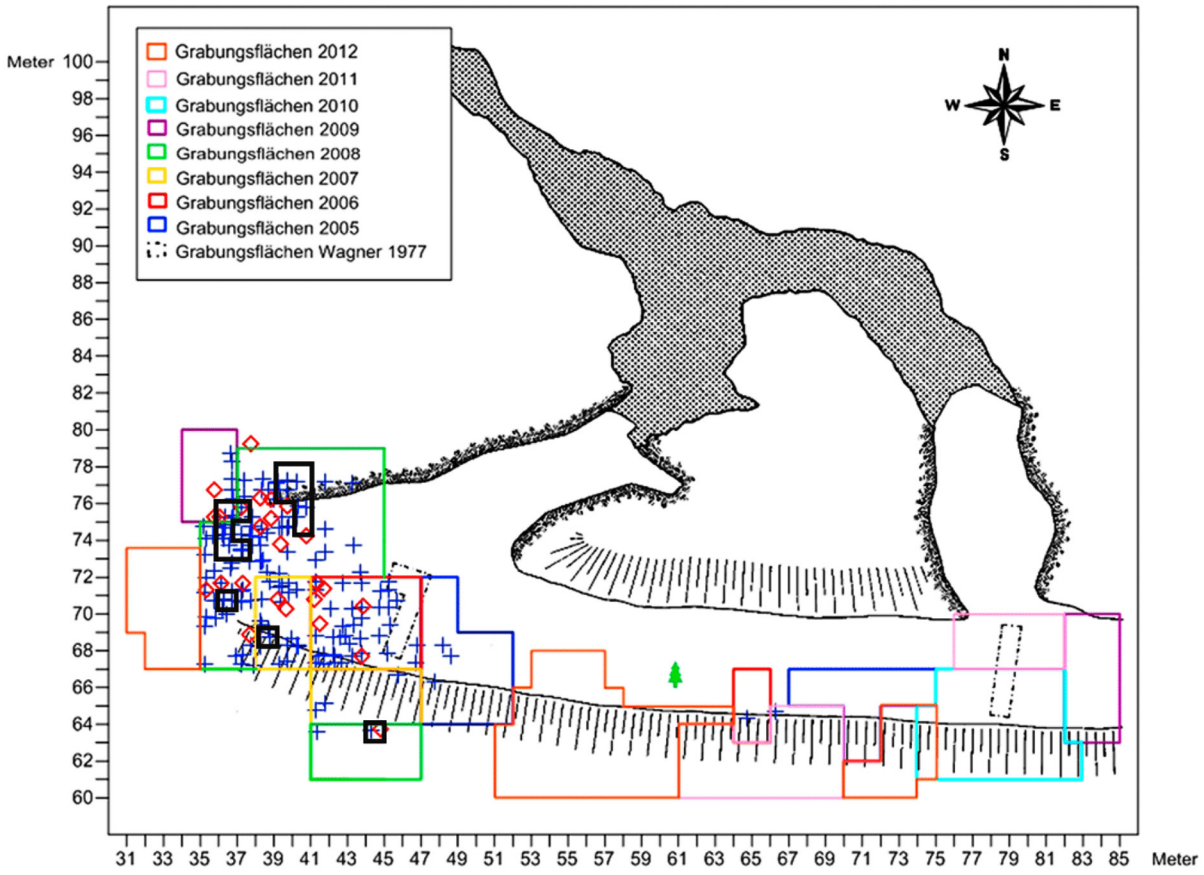


Figure S1. The plot of ZooMS sampled locations in black boxes (image modified from Wolf and Conard 2015). All double-perforated (blue cross) and single-perforated (red lozenge) beads were discovered in front of the southwest entrance, Vogelherd Cave.

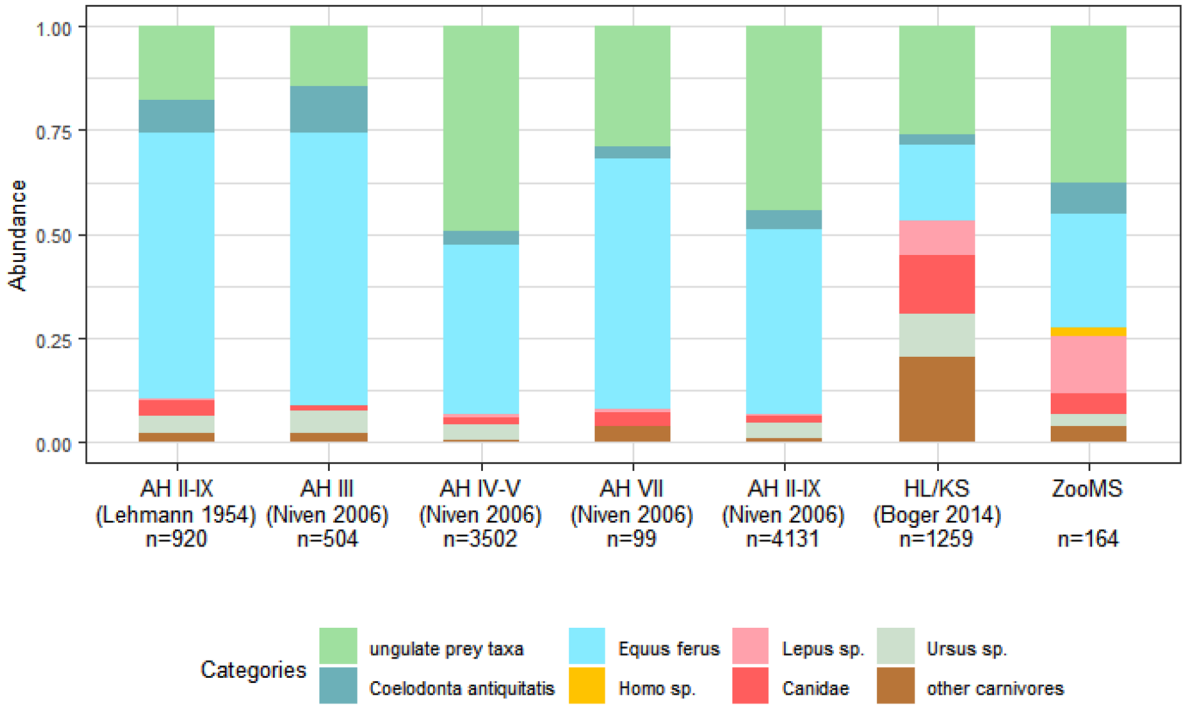


Figure S2. Abundance of identifiable specimens by taxonomic categories, using NISP values, Elephantidea excluded. Data from Boger et al. (2014), Lehmann (1954), and Niven (2006).

SUPPLEMENTARY REFERENCES

Buckley, M., Collins, M., Thomas-Oates, J., Wilson, J.C., 2009. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Comms. Mass Spectrom. 23, 3843–3854.

Buckley, M., Whitcher Kansa, S., 2011. Collagen fingerprinting of archaeological bone and teeth remains from Domuztepe, South Eastern Turkey. Archaeol. Anthropol. Sci. 3(3), 271–280.

Janzen, A., Richter, K.K., Mwebi, O., Brown, S., Onduso, V., Gatwiri, F., Ndiema, E., et al., 2021. Distinguishing African bovids using Zooarchaeology by Mass Spectrometry (ZooMS): new peptide markers and insights into Iron Age economies in Zambia. PloS One 16(5), e0251061.

Strohalm, M., Kavan, D., Novák, P., Volný, M., Havlícek, V., 2010. mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal. Chem. 82(11), 4648–4651.

Team, R. Core, and Others., 2013. R: A Language and Environment for Statistical Computing. http://r.meteo.uni.wroc.pl/web/packages/dplR/vignettes/intro-dplR.pdf

Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., et al., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte Du Renne. Proc. Nat. Acad. Sci. U.S.A. 113 (40), 11162–11167.

Wickham, H., 2016. ggplot2-elegant graphics for data analysis. Springer International Publishing, Cham, Switzerland.

Wolf, S., Conard, N.J., 2015. Personal ornaments of the Swabian Aurignacian. Palethnologie 7 (December). https://doi.org/10.4000/palethnologie.897

Appendix 2 - Other publication

Yang Xu, Naihui Wang, Shizhu Gao, Chunxiang Li, Pengcheng Ma, Shasha Yang, Hai Jiang, Shoujin Shi, Yanhua Wu, Quanchao Zhang, Yinqiu Cui. "Solving the two-decades-old murder case through joint application of ZooMS and ancient DNA approaches." *International Journal of Legal Medicine* 137, no. 2 (2023): 319-327.

International Journal of Legal Medicine (2023) 137:319–327
https://doi.org/10.1007/s00414-022-02944-5

CASE REPORT

Solving the two-decades-old murder case through joint application of ZooMS and ancient DNA approaches

Yang Xu[1] · Naihui Wang[2] · Shizhu Gao[3] · Chunxiang Li[1] · Pengcheng Ma[1] · Shasha Yang[1] · Hai Jiang[4] · Shoujin Shi[5] · Yanhua Wu[6] · Quanchao Zhang[7,8] · Yinqiu Cui[1,7]

Received: 20 October 2022 / Accepted: 29 December 2022 / Published online: 10 January 2023

Abstract

Bones are one of the most common biological types of evidence in forensic cases. Discriminating human bones from irrelevant species is important for the identification of victims; however, the highly degraded bones could be undiagnostic morphologically and difficult to analyze with standard DNA profiling approaches. The same challenge also exists in archaeological studies. Here, we present an initial study of an analytical strategy that involves zooarchaeology by mass spectrometry (ZooMS) and ancient DNA methods. Through the combined strategy, we managed to identify the only biological evidence of a two-decades-old murder case — a small piece of human bone out of 19 bone fragments — and confirmed the kinship between the victim and the putative parents through joint application of next-generation sequencing (NGS) and Sanger sequencing methods. ZooMS effectively screened out the target human bone while ancient DNA methods improve the DNA yields. The combined strategy in this case outperforms the standard DNA profiling approach with shorter time, less cost, as well as higher reliability for the genetic identification results.

Highlights

• The first application of zooarchaeology by mass spectrometry technique in the forensic case for screening out human bones from bone fragment mixtures.
• Application of ancient DNA technique to recover the highly degraded DNA sequence from the challenging sample that failed standard DNA profiling approaches.
• A fast, sensitive, and low-cost strategy that combines the strengths of protein analysis and DNA analysis for kinship identification in forensic research.

Keywords Kinship identification · Cold-case investigation · Zooarchaeology by mass spectrometry · Ancient DNA · Whole-genome sequencing

✉ Yinqiu Cui
cuiyq@jlu.edu.cn

[1] School of Life Sciences, Jilin University, Changchun 130012, China

[2] Max Planck Institute for the Science of Human History, 07745 Jena, Germany

[3] School of Pharmaceutical Sciences, Jilin University, Changchun 130021, China

[4] Criminal Police Detachment, Qingdao Municipal Public Security Bureau, Qingdao 266034, China

[5] Criminal Investigation Team, Jimo Branch, Qingdao Municipal Public Security Bureau, Qingdao 266205, China

[6] Division of Clinical Research, First Hospital of Jilin University, Changchun 130021, China

[7] Bioarchaeology Laboratory, Jilin University, Changchun 130012, China

[8] School of Archaeology, Jilin University, Changchun 130012, China

Introduction

Bioanalytical chemistry has played a more important role in the field of forensic research since the first introduction of DNA profiling in the 1980s [1]. The development of strategies which are rapid, low-cost, and sensitive for challenging samples will undoubtedly be the trend of future molecular forensic research [2, 3]. And various techniques have been continuously upgraded for identifying different biological materials in forensic investigations. In recent years, proteomic evidence (mainly from bodily fluids and skin remains) has also been used in the criminal justice community [4–6]. The rapid development of mass spectrometry technology allows trace amount protein/peptide determination, showing great potential in forensic practice.

Obtaining robust evidence from severely degraded skeleton remains is still a major challenge in forensic practices, especially when the sample is a mixture of undiagnostic bone fragments. With DNA analysis only, more effort is needed on taxonomical identification than on biometrics recognition of individual. However, this kind of samples has commonality with archaeological bone materials, in terms of the specimen's preservation and mixed property. It is worth referring to archaeological approaches for a high-efficient forensic solution. Recent advances in ancient DNA technology, particularly the application of next-generation sequencing (NGS), enable us to recover severely damaged DNA sequence from even ancient samples; thus, it could become an applicable tool for genetic analysis in forensic cases [7, 8].

Zooarchaeology by mass spectrometry (ZooMS) is a proteomic approach based on the collagen peptide mass fingerprinting (PMF) technique, providing taxonomic information through the detection of tryptic peptides of two type I collagens via matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Collagen is phylogenetically informative and it can persist for longer periods than DNA [9–11]. Since invented [12], ZooMS has been used in diverse research fields including archaeology and paleontology, ecology and conservation, as well as cultural heritages [13], applied on a wide range of collagenous materials such as leather, ivory, and parchment [14–17], but mostly on bones. And the technique lends itself particularly well to being utilized for the large-scale taxonomic investigations of faunal assemblages as well as identification of animal remains or products lacking diagnostic features for traditional zooarchaeological determinations. Compared to the DNA taxonomic approaches, ZooMS has advantages including simple procedures, high-throughput, and low detection cost. Previous studies have shown that this technique could provide a quite high identification success rate (>95%) for the analysis of archaeological samples from Late Pleistocene [18–21]. The ZooMS method could no doubt provide reliable taxonomic information for mixed forensic samples before performing the DNA profiling analysis.

In this case, we creatively combined two mature techniques from different fields to solve a cold case that seems to have reached a dead end. Firstly, the ZooMS method is performed to screen out human bones from a mixture of bone fragments, with low cost and short time; then ancient DNA technology is used to recover the highly degraded DNA sequence. Following the analysis of the ancient genome to assess the sex and probable geographic origins, together with the kinship possibility calculation between the individual and the putative parents, we finally managed to provide vital biological evidence for the case.

Case history

In 2002, a 9-year-old boy in Qingdao, Shandong Province, went missing after school. The parents had searched across China for years in vain. After 20 years, the police finally locked the suspect, who had confessed guilty of murdering the boy, burying the body in his own yard filled with domestic wastes (including animal bones), and moved the body to the cropland which is unable to be located several years after the murder. Therefore, efforts to identify the victim's remains ran into difficulties; only a few undiagnostic bone fragments were found in the yard deposit and most of them are less than 2 cm in size (too small to be identified morphologically) and extremely porous and fragile. Routine forensic identification approaches including STR testing was conducted on some of the bones but failed to retrieve any valuable genetic information, probably due to the high degradation level of the DNA molecules in the samples. Since it is of great urgency to identify whether there was any bone(s) that belongs to the victim, a mixture of 19 bone fragments was finally transferred to the ancient DNA laboratory at Jilin University.

Materials and methods

Samples

The responsible justice department provides all the samples and authorizes all protein and DNA testing. Nineteen bone samples were collected and photographed, as shown in Fig. 1. The bones were small, from 0.9 to 4.4 cm in length, bearing no morphological feature. The surface of some bones was damaged by postmortem erosion in the humic soil. To remove any contaminants attached to the surface of the bone, a small area was

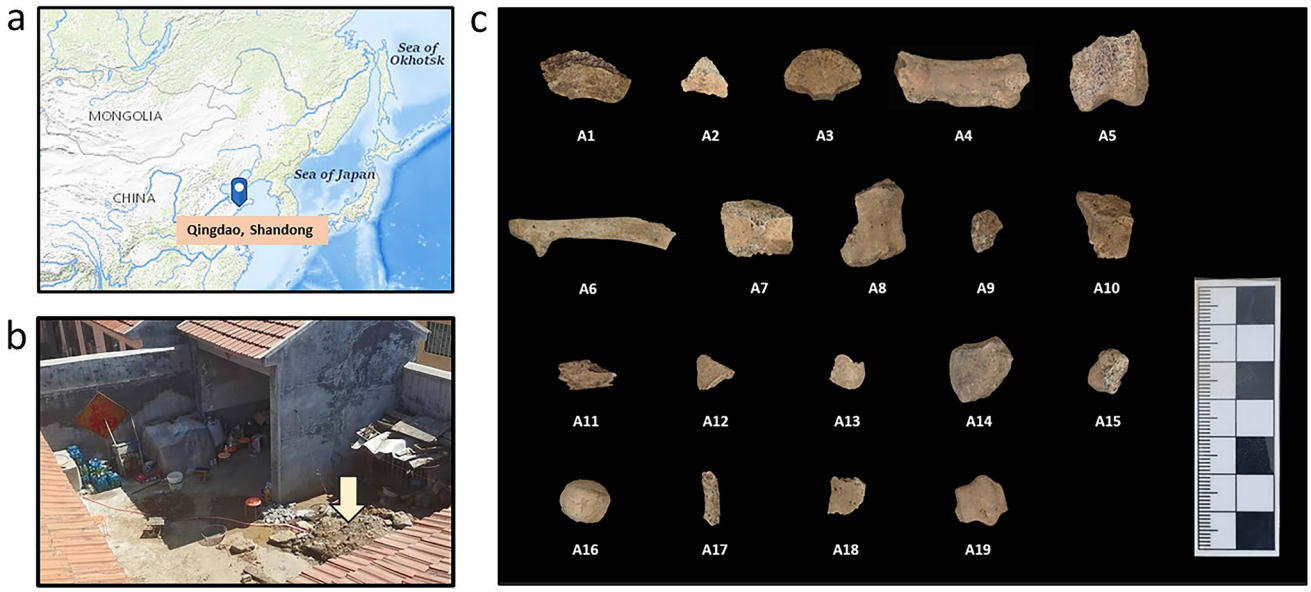


Fig. 1 Location of the burial site and the bone fragments analyzed in this study. **a** Location of Qingdao City, where the murder case took place. The base map was obtained from the USGS National Map Viewer, public domain (http://viewer.nationalmap.gov/viewer/). **b** The burial site in the suspect's yard. **c** Nineteen bone fragment samples excavated at the burial site, and the test numbers are marked at the bottom of each sample

sandblasted. Around 50 mg of bone chip was sub-sampled from each sample for ZooMS. The amount for the following DNA analysis is also about 50 mg, in powder. In our study, the blood samples from the putative mother and father (pM and pF) were also collected on the FTA cards as references.

Taxonomic identification by ZooMS

The collagen in the bone samples was first extracted following the established acid-insoluble protocols [22]: samples were demineralized in 500 µL 0.5 M HCl for 6 h at 4 °C until the bone chips became spongy. The supernatant was then removed and the chips were rinsed 3 times using 0.5 M NH_4HCO_3 until a neutral pH was reached. The chips were incubated at 65 °C for 1 h, in 100 µL of 50 mM NH_4HCO_3. Following incubation, 50 µL supernatant was collected and was digested with 0.5 µg trypsin at 37 °C for 18 h. One microliter of 5% trifluoroacetic acid (TFA) was added to end the digestion. The digested samples were concentrated and desalted using C18 ZipTips, then washed with 200 µL 0.1% TFA and eluted with 50 µL 50% ACN/0.1% TFA (v/v). One blank was analyzed alongside samples as a negative control. Then MALDI-TOF-MS analysis was carried out to obtain the collagen peptide mass spectra for all samples (Supplementary Text), and species were identified using previously published type I collagen peptide markers from reference spectra [23].

DNA analyzing

The only bone fragment identified as human bone by ZooMS (A14, see below) was then conducted by DNA analysis. The genetic relationship identification by routine STR test failed, which might be due to the high degradation level of the DNA molecules in the bone fragment. Therefore, strict ancient DNA protocols were applied during DNA exaction, pre-PCR, and DNA library construction of the bone sample [24], while DNA extraction and library preparation for blood samples from pM and pF were undertaken in the modern genetic laboratory. And both shotgun sequencing and Sanger sequencing were performed on all three samples.

DNA extraction, library preparation, and shotgun sequencing

DNA extraction of A14 was performed in a laboratory for ancient DNA located in the College of Archaeology, Jilin University, and treated as described by Li et al. [25] (Supplementary Text). DNA extraction and library preparation for samples from pM and pF were undertaken by TruePrep® Flexible DNA Library Prep Kit for Illumina (Vazyme, China) in the modern genetic laboratory in the School of Life Sciences, Jilin University, and all the experiments were performed only after genome profile was obtained for bone fragment to avoid the possibility of contaminating the "cold case" remains with the modern reference DNA sample. All the libraries were sent to sequence on Illumina HiSeq X Ten platform.

Authenticity control

In order to evaluate possible contamination of A14 and verify the authenticity of the results, we computed the proportion of C-to-T deamination errors at both the 5′ and 3′ ends of the sequencing reads to evaluate the postmortem damage patterns and then examined mtDNA contamination using Schmutzi [26], an approach that calls endogenous DNA based on the deamination patterns and computes the contamination rate by comparison to a set of known contaminants.

Sanger sequencing

With over 99% accuracy, the Sanger sequencing method remains the "gold standard" for individual identification or kinship testing. To validate mtDNA variants that firstly identified through NGS, we further perform Sanger sequencing targeting the mtDNA hypervariable region I (HVR-I). Two sets of overlapping primers were used to amplify the mtDNA HVR-I between positions 16035 and 16409, and PCR amplifications were done for A14 as described by Li et al. [25], but increasing the number of PCR cycles to 40. Amplification products were sequenced directly using the Sanger sequencing method (ABI PRISM 3130). The amplification and sequencing from bone extract were repeated twice. Extraction blanks and PCR negative controls were carried out for each PCR experiment.

Genomic data processing

The Sanger sequencing result was converted to mtDNA sequence information by Chromas (http://technelysium.com.au/wp/chromas/); the consensus mtDNA sequence obtained from multiple overlapping PCR amplifications was compared to those obtained from pM and pF. For the processing of the shotgun results, the raw fastq files from Illumina platform were processed in EAGER v1.92.50 program [27], an automated computational pipeline specially designed for ancient DNA data processing, which is described in detail in Supplementary Text. The biological sex of A14 was assessed by computing the ratio of X chromosome derived shotgun sequencing data to the autosomal coverage. We measured the rate of damage using mapDamage v2.0.6 [28].

Genetic structure analysis

The uniparental haplogroups of A14, pM, and pF were assigned, and the procedure is described in Supplementary Text. Briefly, the mtDNA consensus sequences were generated using the Geneious software [29], and then determined their mtDNA haplogroups using HaploGrep2 [30]. The male Y chromosome haplogroup was determined by examining a set of positions on the 25,660 diagnostic positions of the ISOGG database (https://isogg.org/), and assigned the final haplogroups by the most downstream derived SNPs. The whole-genome data of three samples in this case was compared to modern populations in the Affymetrix Human Origins (HO) public dataset [31, 32] or the high-coverage Simons Genome Diversity Project [33, 34] and the final dataset consists of 593,124 autosomal SNPs. The genetic affinities of our samples with present-day Asian populations were assessed by principal component analysis (PCA) and outgroup f3 statistics (see Supplementary Text for more details). The PCA was carried out using the "lsqproject" options in the smartpca program [35]. We also implemented outgroup f3 statistics using qp3Pop (v435) program in the ADMIXTOOLS package [36] with Mbuti population from Central Africa as an outgroup, and the f3-statistics were performed on the 1240k dataset.

Genetic relatedness estimation between A14 and the putative parents

Because of the relative low coverage of DNA data from the bone fragment, we applied pairwise mismatch rate (PMR) methods to determine the genetic kinship between A14 and his putative relatives [37]. The PMR approach was designed specially to estimate kinship of ancient samples, by calculating the pairwise mismatch rate of haploid genotypes across autosomal SNPs. The PMR value for each pair of individuals was defined by dividing the number of SNP sites for which two individuals have different alleles sampled by the total number of sites covered in both individuals. In general, the PMR of the identical individuals ($r = 1$) should be half of that between the unrelated individuals ($r = 0$, identified as the population baseline, no inbreeding). Likewise, the PMR for first- ($r = 0.5$) and second-degree relatives ($r = 0.25$) should be 3/4 and 7/8 of the baseline, and more details were described in Supplementary Text.

Results

ZooMS results

After comparing the spectra generated from MALDI-TOF-MS against the published reference [18, 23], taxonomic information of the samples was obtained. Eighteen out of the 19 samples were identified and fortunately, one of the samples, A14, was identified as human (*Homo sapiens*) (Fig. 2). Among the rest 17 samples, one was an avian sample (A6) which could be a chicken or a duck, and the other 16 samples were all identified as pig (*Sus scrofa*). The results are consistent with the fact that they were from a domestic waste deposit. For sample A11, only 3 markers (*m/z* 1105, *m/z* 1453,

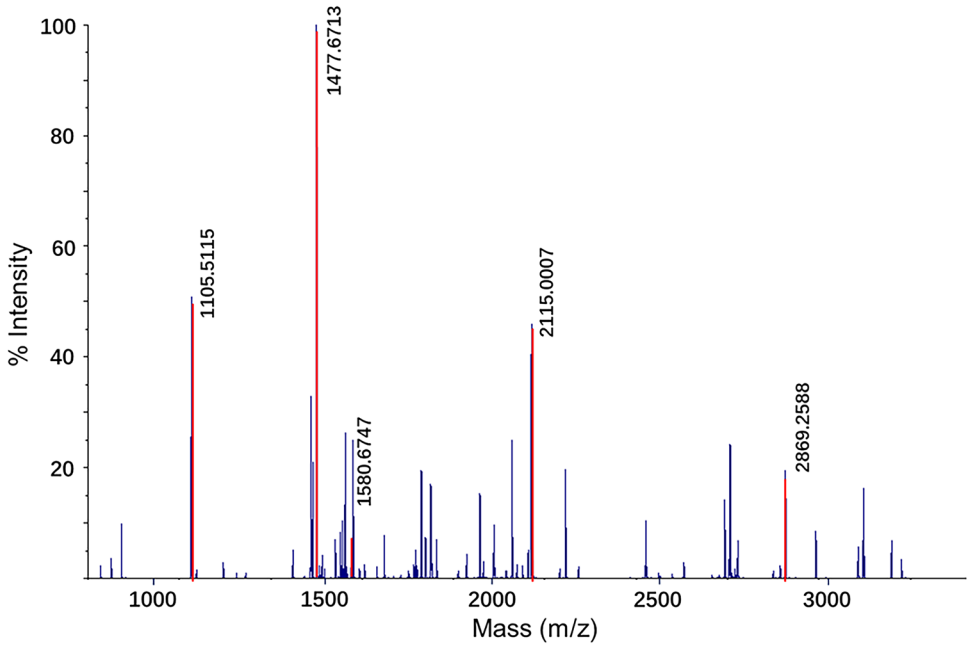


Fig. 2 Spectrum of A14 obtained by the MALDI-TOF mass spectrometer. Five peptide markers assigned A14 to *Homo sapiens* are colored in red

and *m/z* 2820) were detected, which is not enough for more accurate taxonomic identification, indicating a higher level of degradation. The blank control returned negative result, and no cross-contamination was observed. Details of the ZooMS results are shown in Supplementary Fig. S1 and Table S1. The ZooMS taxonomic identification was completed in 3 days, with the cost less than 10 dollars per sample.

DNA analysis results

The authenticity of the genome data

We applied strict procedures to minimize exogenous DNA contamination following the ancient DNA standard. During the experiment process, the negative extraction and amplification controls were free of contamination, and the multiple sequencing results were consistent, including twice Sanger sequencing and shotgun sequencing. Through analyzing the characteristics of genomic library reads, we also observed that the A14 exhibited postmortem chemical damage signatures of DNA molecular, such as small average sequence size with 75 bp, and the nucleotide misincorporation patterns at the 3′- and 5′- ends of the DNA sequences (Table 1, Supplementary Fig. S2). Meanwhile, the sequence reads from A14 showed a low level of contamination for mtDNA (0.5%). The results proved our previous degradation assessment of the sample, and verified the authenticity of A14 data as well. Our sex determination results also show the A14 was from a male individual.

Genetic analyses of A14

We successfully extract endogenous DNA from A14 and the DNA library was sequenced to a low coverage with 0.044×. To characterize the genetic profile of the A14, we implemented principal component analysis (PCA) of present-day Asian people and A14 genome. The results show that the genetic distributions of modern people are consistent with the geographic locations in the PCA plot [38]. We found that the A14 is falling in the group of modern Han and clustered with the putative parents (Fig. 3b). The observation from the PCA plot was further confirmed by the outgroup f3 statistics in the form of (Mbuti; X, A14), where X was represented by worldwide populations; the result showed significant allele sharing between A14 and the putative parents, followed by Eastern Asia populations, such

Table 1 High-throughput sequencing result of DNA library of A14, pM, and pF

Sample name	Endogenous DNA (%)	Mean coverage	SNP	Average fragment length	mt-hg	Y-hg	mt-contamination
A14	19.776	0.0548	50021	75.08	D4j3	O2a2b1a1a5	0.005
pM	94.763	1.0067	597100	163.11	D4j3	-	-
pF	94.272	0.8785	546202	170.4	D4g2a1	O2a2b1a1a5	-

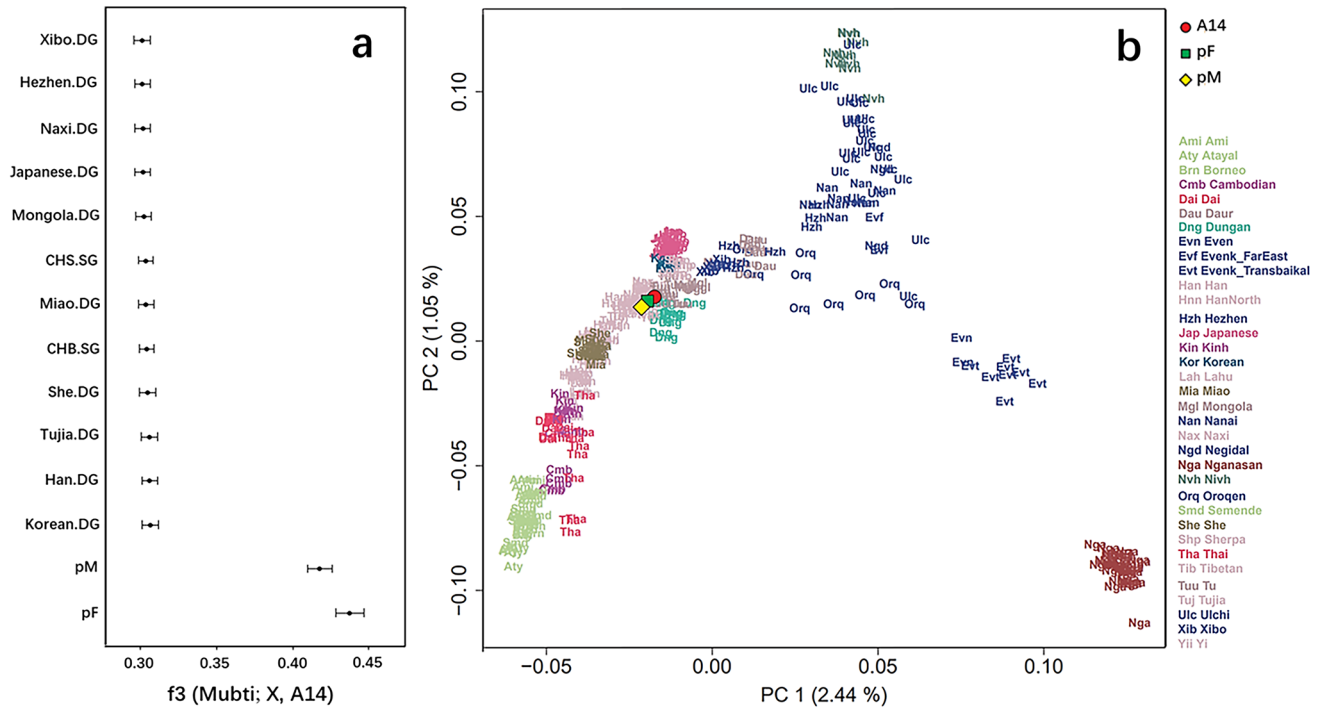


Fig. 3 The genetic profile of A14 by outgroup f3 and PCA. **a** The top 14 populations (including the putative parents) sharing the highest amount of genetic drift with A14 measured by f3 (Mbuti; X, A14). **b** Principal component analysis of A14 and the putative parents projected onto present-day Asian populations

as Han, Korean, and Tujia (Fig. 3a), indicating a much closer genetic affinity of A14 with the putative parents.

Uniparental and autosomal genetic kinship analysis with putative parents

We retrieved almost complete mtDNA sequence (99.5%) for A14, with an average coverage of 10.16-fold, which were further assigned to an explicit haplogroup of D4j3 which is the most prevalent haplogroup in the modern East Asian populations [39]. To confirm the results of NGS, we used two sets of overlapping primers to amplify the HVR of the mtDNA control region, and used the Sanger sequencing method to obtain the HVR-I sequence, which contains the mutant motif 16184-16223-16311-16362. Sanger sequencing results were consistent with the mitochondrial genome obtained by NGS. The mtDNA genome reconstructed from the shotgun genome data of pM produced the same mtDNA profile as A14. But a different profile was obtained from pF, which belonged to haplogroup D4g2a1 (Table 1, Supplementary Table S2). In contrast, the A14 was assigned to Y chromosome haplogroup O2a2b1a1a5 the same as pF (Table 1, Supplementary Table S3), which is a widespread lineage in modern Northeastern Asians such as Sino-Tibetan-speaking populations [40]. Uniparental results showed probable maternal and paternal kinship between A14 and pM, as well as A14 and pF, independently.

To estimate the genetic relatedness between A14 and the two putative parental samples at a finer scale, the degree of genetic relatedness between individuals from autosomes was determined. We calculated PMR from haplotype genotyping of the "1240k" panel using a special method designed for ancient DNA [41]. The overlapping SNPs pairs between A14 and the test samples reached over 20,000, which is sufficient to avoid the artificial bias caused by the high deletion rate of A14. The PMR value between pM and pF was 0.238, which is similar to the value from unrelated pairs of modern northern Han. This suggested they had no close relatedness with each other, which is consistent with their self-reported genetic background. We therefore treated it as the baseline value, together with those obtained between pairs of unrelated modern Han. The PMR values for A14-pM and A14-pF were 0.178 and 0.176, respectively, roughly 1/2 of the baseline value (Fig. 4), suggesting that A14 shares first-degree relatedness with them.

Discussion

The retrieved genetic profile of A14

Through the joint application of ZooMS and ancient DNA approaches, we efficiently screen out the only human bone, A14, among 19 bone fragments, and retrieve the genetic

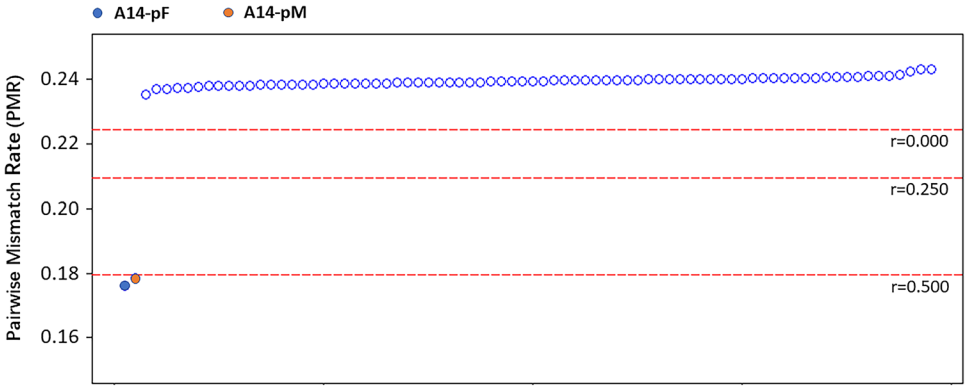


Fig. 4 Pairwise mismatch rate (PMR) between A14 and the putative parents. Baseline was generated by PMR calculated within northern Han individuals ($n = 10$)

profile of the target sample, as follows: the sample is from a human male skeleton, both the uniparental haplogroup distribution and the genome composition of A14 indicated a geographic origin in East Asia together with pM and pF. From uniparental genetic analyses, we found that A14 was potentially maternally related to pM, and paternally related to pF. Autosomal analysis revealed first-degree relatedness of A14 with these two individuals separately, and the genetic data of the two putative parents proved their unrelatedness, from maternal and genomic perspectives, a result consistent with the de facto relationship between the two individuals. Altogether, our combined test results confirmed that the remains A14 most likely belong to the missing 9-year-old boy.

Combined ZooMS and ancient DNA as a promising approach in the forensic practice

Bones are one of the most common biological types of evidence in forensic cases. Due to the preservation conditions or special underground environment, sometimes only a mixture of poorly preserved bones is recovered at the crime scene. The specimens could be fragmentary or lacking diagnostic morphological features. In practice, doing standard DNA profiling on every sample in the mixture is possible but not the most efficient solution. ZooMS, as a taxonomic method, is able to identify the crime-relevant species (like *Homo sapiens*), from lots of irrelevant remains (like bones belonging to *Sus scrofa* and Aves in this case). Another difficulty is that the DNA molecules maybe severely degraded because of the humic environment, resulting in an average length of DNA fragments less than 80 bp. The length is similar to the archaeological samples, making it difficult to perform routine STR tests and other forensic methods commonly used for genetic identification.

In this case, our aim was to obtain the biological evidence of a victim. The bones ($n = 19$) in mixture were highly fragmented and degraded. We designed a two-step analytical strategy in our study. ZooMS is used in the first step to screen the human bone fragments from the mixture, and the whole process only took three workdays. With a relatively low cost, one human bone (A14) was successfully identified. The second step involved profiling methods from the ancient DNA field, and combining the reliable accuracy of Sanger sequencing with the high-throughput nature of NGS, we obtained authentic genomic sequences of the victim, which passed the uniparental genetic marker analysis and the relativeness calculation, matched to the putative parents.

To sum up, through introducing the state-of-the-art technologies in archaeology into the forensic practice, we identified and confirmed the bone fragment of the victim boy with high efficiency. This study provides new evidence for solving cold cases and highlights the enormous potential of multidisciplinary techniques applied in forensic study for crime solving and justice.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00414-022-02944-5.

Author contribution Yang Xu and Naihui Wang contributed equally to this work. Yinqiu Cui and Yang Xu designed the research; Yang Xu, Naihui Wang, and Chunxiang Li performed the experimental work; Yang Xu, Shizhu Gao, Pengcheng Ma, ShashaYang, and Yanhua Wu analyzed data; Hai Jiang and Shoujin Shi provided samples; Quanchao Zhang coordinated with the Qingdao police; and Yang Xu, Naihui Wang, and Yinqiu Cui wrote the paper. All authors read and approved the final manuscript.

Funding This work received supports from the Science and Technology Development Project of Jilin Province (grant numbers 20210508040RQ and 20200201138JC) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 715069-FINDER-ERC-2016-STG).

Declarations

Ethics approval and consent to participate This study does not need ethical approval.

Informed consent Not required.

Conflict of interest The authors declare no competing interests.

References

1. Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318:577–579. https://doi.org/10.1038/318577a0
2. Butler JM (2015) The future of forensic DNA analysis. Philos Trans R Soc B 37:180–195. https://doi.org/10.1098/rstb.2014.0252
3. Kaur L, Sharma SG (2022) Forensic DNA analysis: a powerful investigative tool. In: Singh J, Sharma NR (eds) Crime scene management within forensic science-forensic techniques for criminal. Springer, Singapore
4. Steendam KV, Ceuleneer MD, Dhaenens M, Hoofstat DV, Deforce D (2013) Mass spectrometry-based proteomics as a tool to identify biological matrices in forensic science. Inernational Journal of Legal Medicine 127:287–298. https://doi.org/10.1007/s00414-012-0747-x
5. Akihisa I, Yusuke D, Koichi S (2015) Identification and evaluation of potential forensic marker proteins in vaginal fluid by liquid chromatography/mass spectrometry. Anal Bioanal Chem 407:7135–7144. https://doi.org/10.1007/s00216-015-8877-x
6. Kamanna S, Henry J, Voelcker N, Linacre A, Kirkbride KP (2018) A complementary forensic 'proteo-genomic' approach for the direct identification of biological fluid traces under fingernails. Anal Bioanal Chem 410:6165–6175. https://doi.org/10.1007/s00216-018-1223-3
7. Hofreiter M, Sneberger J, Pospisek M, Vanek D (2021) Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. Forensic Science International: Genetics 54:102538. https://doi.org/10.1016/j.fsigen.2021.102538
8. Zavala EI, Thomas JT, Sturk-Andreaggi K et al (2022) Ancient DNA methods improve forensic DNA profiling of Korean War and World War II Unknowns. Genes 13:129. https://doi.org/10.3390/genes13010129
9. Buckley M, Larkin N, Collins M (2011) Mammoth and mastodon collagen sequences; survival and utility. Geochim Cosmochim Acta 75:2007–2016. https://doi.org/10.1016/j.gca.2011.01.022
10. Cappellini E, Jensen LJ, Szklarczyk D et al (2012) Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. J Proteome Res 11:917–926. https://doi.org/10.1021/pr200721u
11. Welker F (2018) Palaeoproteomics for human evolution studies. Quat Sci Rev 190:137–147. https://doi.org/10.1016/j.quascirev.2018.04.033
12. Buckley M, Anderung C, Penkman K, Raney BJ, Götherström A, Thomas-Oates J, Collins MJ (2008) Comparing the survival of osteocalcin and mtDNA in archaeological bone from four European sites. J Archaeol Sci 35:1756–1764. https://doi.org/10.1016/j.jas.2007.11.022
13. Richter KK, Codlin MC, Seabrook M, Warinner C (2022) A primer for ZooMS applications in archaeology. Proc Natl Acad Sci USA 119:e2109323119. https://doi.org/10.1073/pnas.2109323119
14. Coutu AN, Whitelaw G, P. le Roux, Sealy. J (2016) Earliest evidence for the ivory trade in Southern Africa: isotopic and ZooMS analysis of seventh–tenth century AD ivory from KwaZulu-Natal. Afr Archaeol Rev 33: 411-435 https://doi.org/10.1007/s10437-016-9232-0.
15. Ebsen JA, Haase K, Larsen R, Sommer DVP, Brandt LØ (2019) Identifying archaeological leather–discussing the potential of grain pattern analysis and zooarchaeology by mass spectrometry (ZooMS) through a case study involving medieval shoe parts from Denmark. J Cult Herit 39:21–31. https://doi.org/10.1016/j.culher.2019.04.008
16. Fiddyment S, Holsinger B, Ruzzier C et al (2015) Animal origin of 13th-century uterine vellum revealed using noninvasive peptide Fingerprinting. Proc Natl Acad Sci USA 112:15066–15071. https://doi.org/10.1073/pnas.1512264112
17. Kirby DP, Buckley M, Promise E, Trauger SA, Holdcraft TR (2013) Identification of collagen-based materials in cultural heritage. Analyst 138:4849–4858. https://doi.org/10.1039/c3an00925d
18. Welker F, Hajdinjak M, Talamo S et al (2016) Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proc Natl Acad Sci USA 113:11162–11167. https://doi.org/10.1073/pnas.1605834113
19. Devièse T, Karavanić I, Comeskey D, Kubiak C, Korlević P, Hajdinjak M, Radović S, Procopio N, Buckley M, Pääbo S, Higham T (2017) Direct dating of Neanderthal remains from the site of Vindija Cave and implications for the Middle to Upper Paleolithic transition. Proc Natl Acad Sci USA 114:10606–10611. https://doi.org/10.1073/pnas.1709235114
20. Brown S, Higham T, Slon V et al (2016) Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Sci Rep 6:23559–23566. https://doi.org/10.1038/srep23559
21. Charlton S, Alexander M, Collins M, Milner N, Mellars P, O'Connell TC, Stevens RE, Craig OE (2016) Finding Britain's last hunter-gatherers: a new biomolecular approach to 'unidentifiable' bone fragments utilising bone collagen. J Archaeol Sci 73:55–61. https://doi.org/10.1016/j.jas.2016.07.014
22. Wang N, Brown S, Ditchfield P et al (2021) Testing the efficacy and comparability of ZooMS protocols on archaeological bone. J Proteome 233:104078–104090. https://doi.org/10.1016/j.jprot.2020.104078
23. 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 Commun Mass Spectrom 23:3843–3854. https://doi.org/10.1002/rcm.4316
24. Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. Science 289:1139. https://doi.org/10.1126/science.289.5482.1139b
25. Li H, Zhao X, Zhao Y, Li C, Si D, Zhou H, Cui Y (2011) Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. J Hum Genet 56:815–822. https://doi.org/10.1038/jhg.2011.102
26. Renaud G, Slon V, Duggan AT, Kelso J (2015) Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol 16:224. https://doi.org/10.1186/s13059-015-0776-0
27. Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K (2016) EAGER: efficient ancient genome reconstruction. Genome Biol 17:60. https://doi.org/10.1186/s13059-016-0918-z
28. Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E, Orlando L (2011) mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27:2153–2155. https://doi.org/10.1093/bioinformatics/btr347
29. Kearse M, Moir R, Wilson A et al (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

30. Weissensteiner H, Pacher D, Kloss-Brandstatter A, Forer L, Specht G, Bandelt HJ, Kronenberg F, Salas A, Schonherr S (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res 44:W58–W63. https://doi.org/10.1093/nar/gkw233
31. Skoglund P, Posth C, Sirak K et al (2016) Genomic insights into the peopling of the Southwest Pacific. Nature 538:510–513. https://doi.org/10.1038/nature19844
32. Lazaridis I, Nadel D, Rollefson G et al (2016) Genomic insights into the origin of farming in the ancient Near East. Nature 536:419–424. https://doi.org/10.1038/nature19310
33. Mallick S, Li H, Lipson M et al (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538:201–206. https://doi.org/10.1038/nature18964
34. Jeong C, Balanovsky O, Lukianova E et al (2019) Characterizing the genetic history of admixture across inner Eurasia. Nat Eco Evol 3:966–976. https://doi.org/10.1038/s41559-019-0878-2
35. Patterson TG, Joseph S (2006) Development of a self-report measure of unconditional positive self-regard. Psychol Psychother 79:557–570. https://doi.org/10.1348/147608305x89414
36. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012) Ancient admixture in human history. Genetics 192:1065–1093. https://doi.org/10.1534/genetics.112.145037
37. Kennett DJ, Plog S, R. J. George, et al. (2017) Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nat Commun 8: 14115. https://doi.org/10.1038/ncomms14115.
38. Ning C, Li T, Wang K et al (2020) Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat Commun 11:2700. https://doi.org/10.1038/s41467-020-16557-2
39. Dryomov SV, Nazhmidenova AM, Shalaurova SA, Morozov IV, Tabarev AV, Starikovskaya EB, Sukernik RI (2015) Mitochondrial genome diversity at the Bering Strait area highlights prehistoric human migrations from Siberia to northern North America. Eur J Hum Genet 23:1399–1404. https://doi.org/10.1038/ejhg.2014.286
40. Yan S, Wang CC, Li H, Li SL, Jin L (2011) An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4. Eur J Hum Genet 19:1013–1015. https://doi.org/10.1038/ejhg.2011.116
41. Mathieson I, Lazaridis I, Rohland N et al (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503. https://doi.org/10.1038/nature16152

International Journal of Legal Medicine

Supplementary Information for

Solving the two-decades-old murder case through joint application of ZooMS and ancient DNA approaches

Yang Xu[a], Naihui Wang[b], Shizhu Gao[c], Chunxiang Li[a], Pengcheng Ma[a], Shasha Yang[a], Hai Jiang[d], Shoujin Shi[e], Yanhua Wu[f], Quanchao Zhang[g,h], Yinqiu Cui[a,g*]

[a] School of Life Sciences, Jilin University, 130012, Changchun, China;
[b] Max Planck Institute for the Science of Human History, 07745, Jena, Germany;
[c] School of Pharmaceutical Sciences, Jilin University, 130021, Changchun, China;
[d] Criminal Police Detachment, Qingdao Municipal Public Security Bureau, 266034, Qingdao, China;
[e] Criminal Investigation Team, Jimo Branch, Qingdao Municipal Public Security Bureau, 266205, Qingdao, China;
[f] Division of Clinical Research, First Hospital of Jilin University, 130021, Changchun, China;
[g] Bioarchaeology Laboratory, Jilin University, 130012, Changchun, China;
[h] School of Archaeology, Jilin University, 130012, Changchun, China.

Corresponding author

* Yinqiu Cui, School of Life Sciences, Jilin University, 130012, Changchun, China, +86 13604337044, cuiyq@jlu.edu.cn

This file includes:

Supplementary Text

Supplementary Figures

Supplementary Tables

Supplementary References

MALDI-TOF-MS analysis

To prepare peptide samples for MS analysis, 0.5 µL of the resulting elution was spotted with an equal volume of α -cyano-4-hydroxycinnamic acid solution (10 mg/mL in 50 % ACN/0.1 % TFA (v/v)) on an Opti-TOF 384 MALDI plate insert (AB SCIEX, USA) prior to the analysis. Each sample or blank were spotted in triplicate. MALDI-TOF-MS analysis was carried out on a 5800 MALDI-TOF/TOF mass spectrometer (AB SCIEX, USA) coupled with a 355-nm Nd-YAG laser. The laser energy was adjusted to 4000, and all MS spectra were acquired in batch using the reflector detection in the positive ionization mode. Calibration was performed using a 6-peptide calibration mixture (Tube PN: 4368762, AB SCIEX, USA) to ensure mass accuracy within m/z 0.1. Mass spectra files were processed using the official data analysis software for MALDI-TOF/TOF mass spectrometer (AB SCIEX, USA) -Data Explorer version 4.3. Processing of raw spectra was conducted in Data Explorer with a peak picking algorithm that used a signal to noise ratio of 70.

DNA extraction, library preparation and shotgun sequencing for A14

DNA extraction of A14: Bone powder (50 mg) were incubated in a 3 mL solution containing 0.45 M ethylene diamine tetraacetic acid, 0.5 % SDS and 0.7 mg mL-1 proteinase K at 50 °C in a shaker (220 rpm/min) for 24 h, DNA was extracted using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Two separate libraries were prepared from 30 µL bone DNA extract as described in our previously published paper [1], except that 1:10 diluted adapter was applied to the ends of DNA fragments during ligation. The quality and concentration of these two libraries and one library negative control were determined on an Agilent Bioanalyzer 2100 and multiplex shotgun sequencing was carried out using Illumina HiSeq X Ten platform.

Shotgun data processing

For the processing of the shotgun results, the raw fastq files from Illumina platform were processed in EAGER v1.92.50 program, an automated computational pipeline specially designed for ancient DNA data processing [2]. Specifically, in EAGER, Illumina Adapters were trimmed from sequencing data with AdapterRemoval v2.2.0 [3] and read length shorter than 30bp were discarded. The trimmed data was then mapped to the human reference genome (GRCh37) using BWA 0.7.12 with '-n 0.01' and '-l 1024' to allow for more mismatches and to disable the seeding. The duplicated reads were then removed with dedup v0.12.2 [2] and sequences with a mapping quality of ≥ 30 are retained using SAMtools [4]. Finally, we randomly called genotype for a SNP from trimmed reads with high-quality base (Q > 30) that implemented using pileupCaller (https://github.com/stschiff/sequenceTools).

Uniparental haplogroup assignment

To determine the mtDNA haplogroup, we first aligned the adapter trimmed reads to the revised Cambridge Reference Sequence (rCRS; NC_012920.1) and removed low-quality sequences (-q30). Next, we generated the mtDNA consensus sequences of our ancient individuals using the Geneious v11.1.3 (https://www.geneious.com/) and then assigned their mtDNA haplogroups using HaploGrep272.
We determined the male Y chromosome haplogroup by examining a set of positions on the 25,660 diagnostic positions on the ISOGG database (https://isogg.org/) and assigned the final haplogroups by the most downstream derived SNPs.

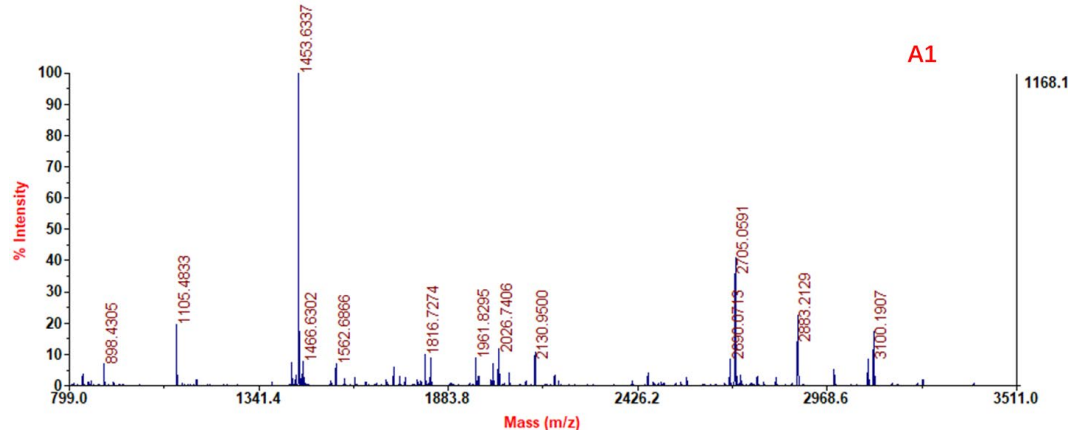
Genetic structure analysis

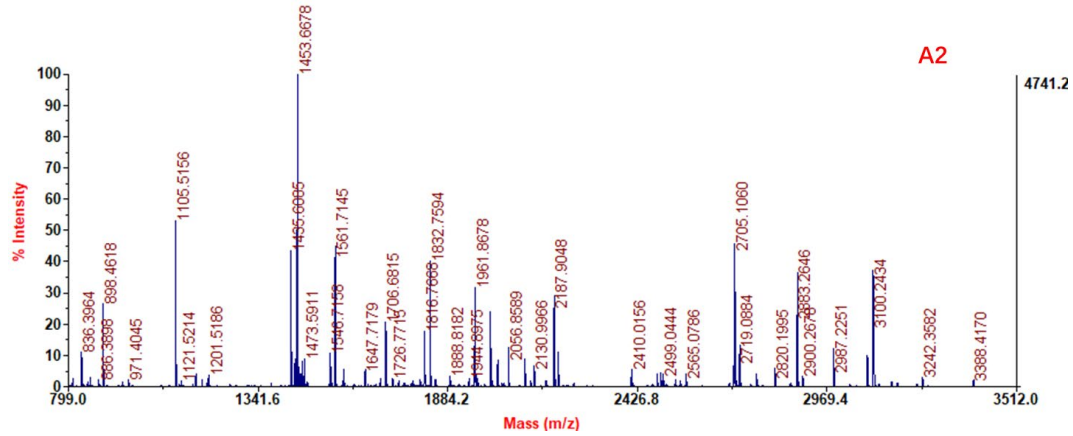
Principal components analysis (PCA), is a statistical method commonly used in population genetics to identify structure in the distribution of genetic variation across geographical location and ethnic background. When PCA was performed to low coverage data such as ancient genomic data, we usually construct the Eigenvectors firstly from the high-quality set of modern samples in the HO set, and then project the ancient or low coverage samples onto these Eigenvectors. This allows our A14 with as few as 50,000 SNPs to project into correct location of PCA plot (compare with ~600,000 SNPs for HO samples). In this case, we performed PCA for the genomic data of three samples as implemented in the smartpca v16000 in the Eigensoft v7.2.1 with default parameters, and Shrinkmode: YES and lsqproject: YES options to minimize bias due to high missing rate.

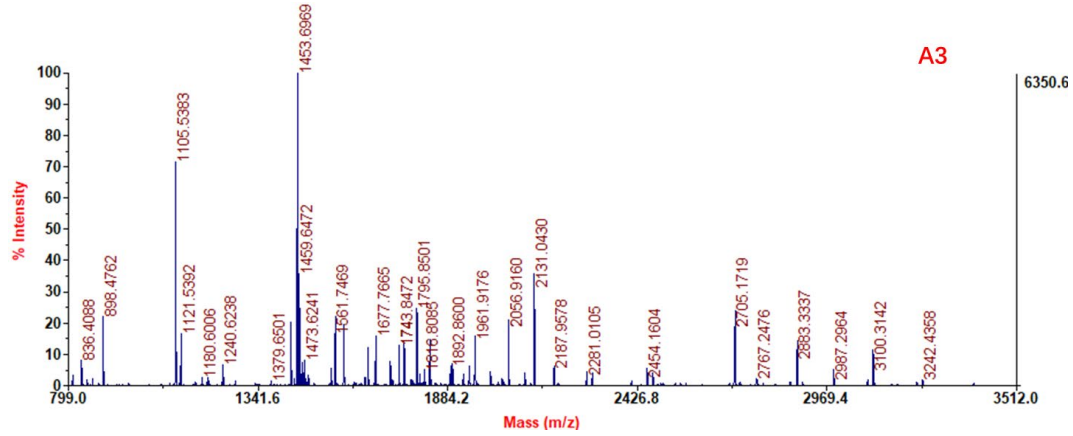
The outgroup f3-statistics measure allele frequency correlations between populations to understand population relationships. In this case, we calculated f3-statistics were using the qp3Pop (v435) programs in the ADMIXTOOLS v5.1 package using default parameters [5].

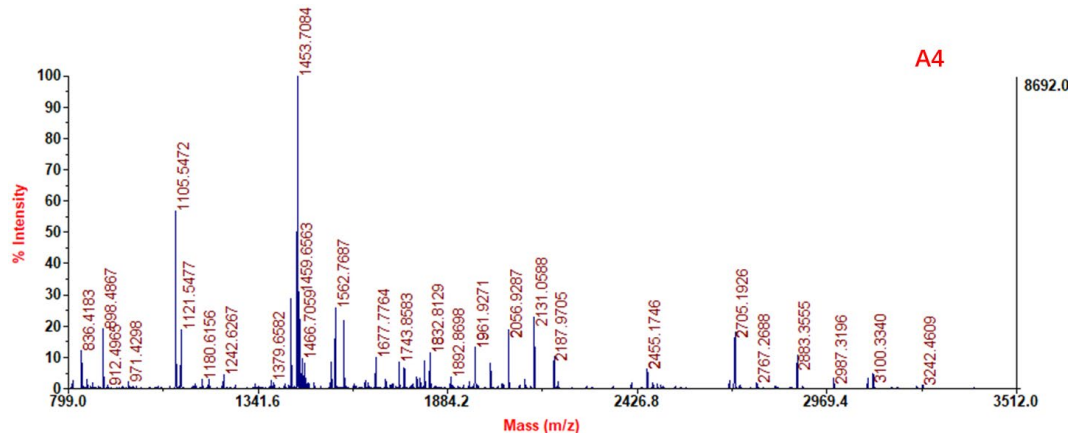
Genetic relatedness analysis

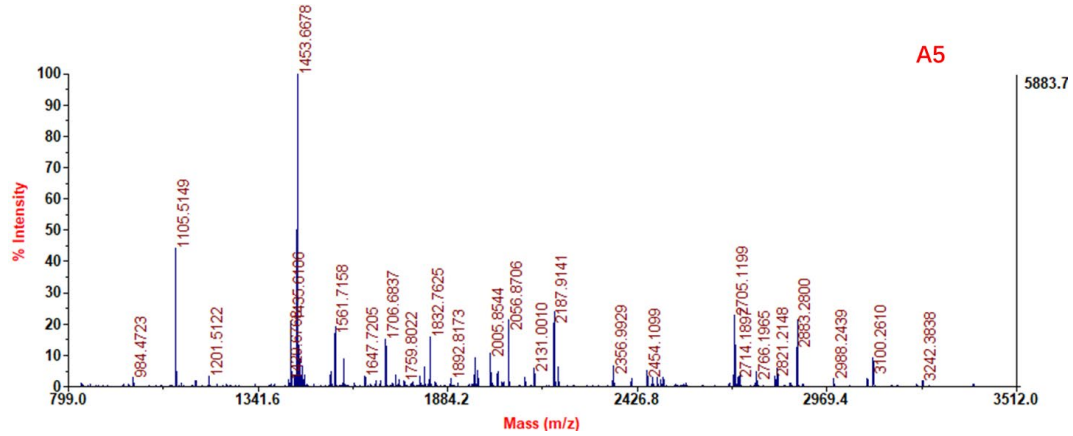
Genetic relatedness was estimated by calculating pairwise mismatch rate between each pair of individuals. The pairwise mismatch rate provides an indication of close genetic relationships, such as identical individuals/twins, first and second degree relatives. In this study, we tested relatedness among our three individuals by calculating PMR of haploid genotypes across autosomal SNPs in the 1240K data set, following the idea present by Kennett et al. [6]. PMR between unrelated individuals has a baseline value, which was obtained by estimating by the empirical distribution of PMR for multiple individuals, for example, ten unrelated modern northern Han individuals in this case. PMR values between A14 and pM, A14 and pF, pM and pF were calculated to finally estimate the genetic relatedness between A14 and the two putative parental samples. Notably, pairwise mismatch rate between two samples from the same individual (r=1) is expected to be a half of that between two unrelated individuals (r=0). Likewise, PMR for the first-degree relative pair (r=0.5) is expected to be three-quarters of the baseline. In general, pairwise mismatch rate is a linear function of the coefficient of relationship. More detailed description of the method can be found in the Supplemental Materials of Jeong et al.[7]

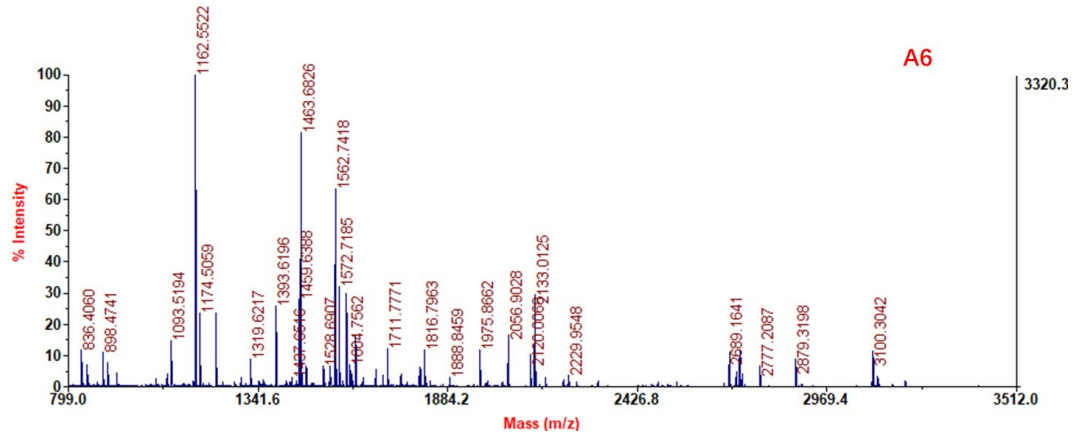

A1
% Intensity
Mass (m/z)
1453.6337
1105.4833
898.4305
1466.6302
1562.6866
1816.7274
1961.8295
2026.7406
2130.9500
2690.0713
2705.0591
2883.2129
3100.1907
1168.1
799.0
1341.4
1883.8
2426.2
2968.6
3511.0

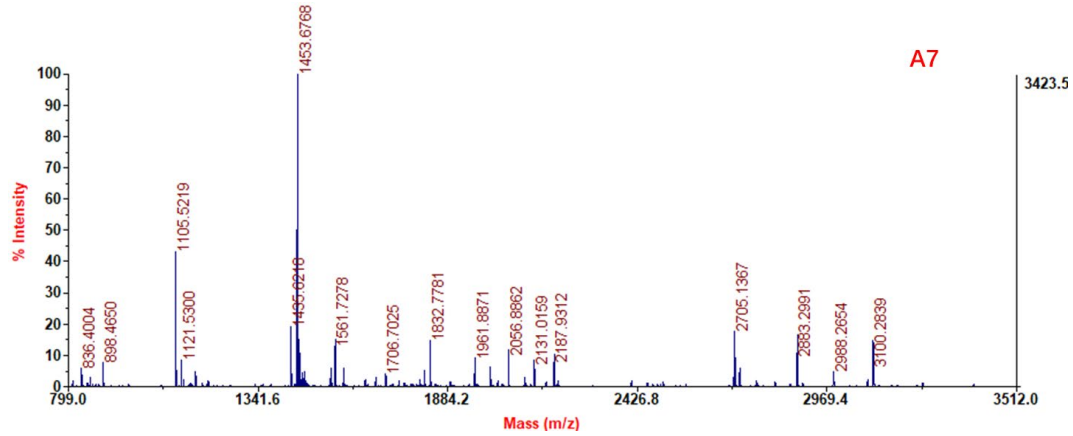

A2
% Intensity
Mass (m/z)
1453.6678
1105.5156
898.4618
836.3964
971.4045
1121.5214
1201.5186
1455.6805
1561.7145
1473.5911
1647.7179
1706.6815
1726.7719
1832.7594
1888.8182
1961.8678
2056.8589
2130.9966
2187.9048
2410.0156
2499.0444
2565.0786
2705.1060
2719.0884
2820.1995
2883.2646
2987.2251
3100.2434
3242.3582
3388.4170
4741.2
799.0
1341.6
1884.2
2426.8
2969.4
3512.0

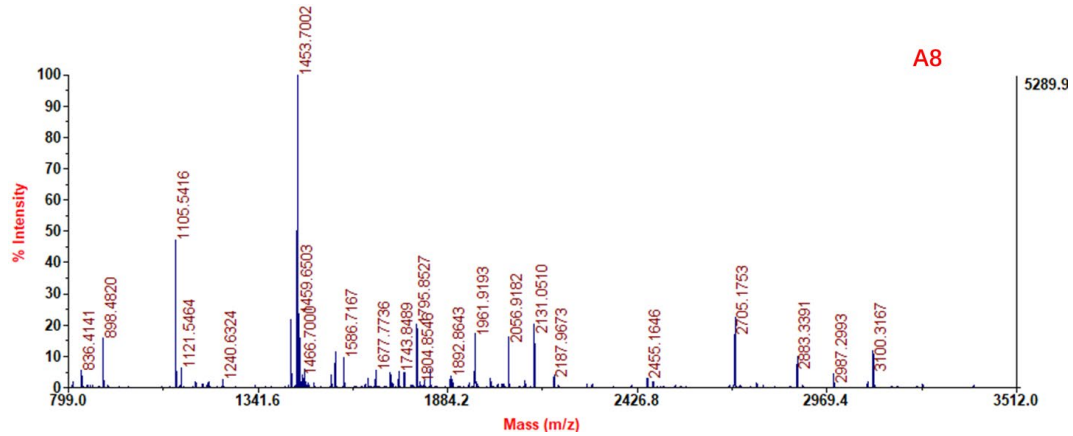

A3
% Intensity
Mass (m/z)
1453.6969
1105.5383
898.4762
836.4088
1121.5392
1180.6006
1240.6238
1379.6501
1459.6472
1473.6241
1561.7469
1677.7665
1743.8472
1795.8501
1816.8083
1892.8600
1961.9176
2056.9160
2131.0430
2187.9578
2281.0105
2454.1604
2705.1719
2767.2476
2883.3337
2987.2964
3100.3142
3242.4358
6350.6
799.0
1341.6
1884.2
2426.8
2969.4
3512.0

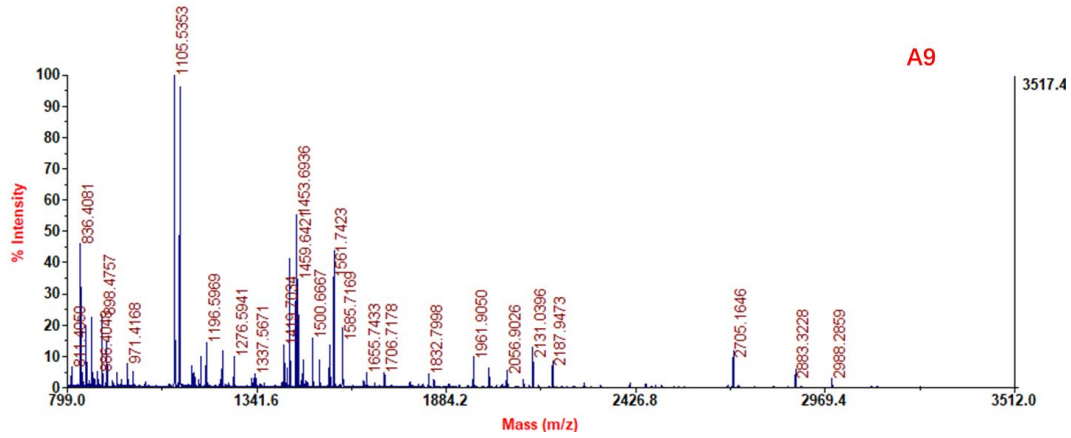

A4
% Intensity
Mass (m/z)
1453.7084
1105.5472
898.4867
836.4183
912.4965
971.4298
1121.5477
1180.6156
1242.6267
1379.6582
1459.6563
1466.7059
1562.7687
1677.7764
1743.8583
1832.8129
1892.8698
1961.9271
2056.9287
2131.0588
2187.9705
2455.1746
2705.1926
2767.2688
2883.3555
2987.3196
3100.3340
3242.4609
8692.0
799.0
1341.6
1884.2
2426.8
2969.4
3512.0

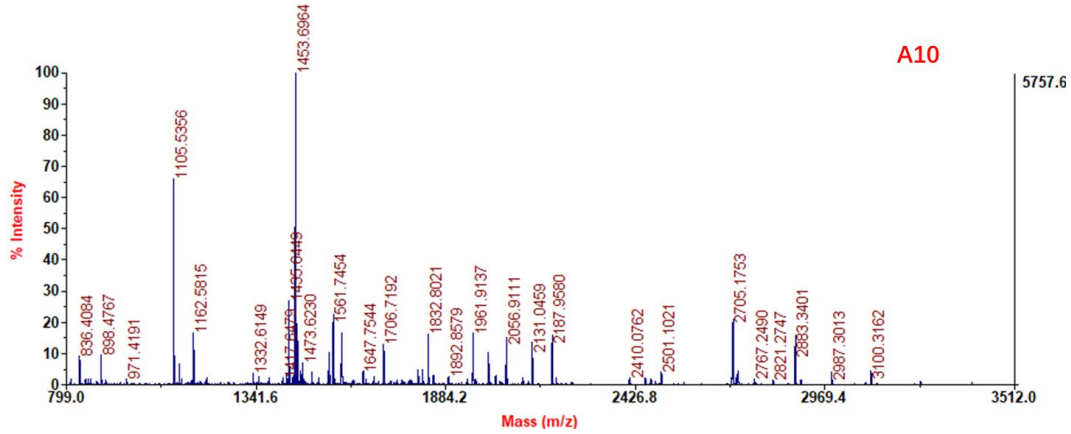

A5
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
5883.7
1453.6678
1105.5149
984.4723
1201.5122
1561.7158
1647.7205
1706.6837
1759.8022
1832.7625
1892.8173
2005.8544
2056.8706
2131.0010
2187.9141
2356.9929
2454.1099
2705.1199
2714.1892
2766.1965
2821.2148
2883.2800
2988.2439
3100.2610
3242.3838

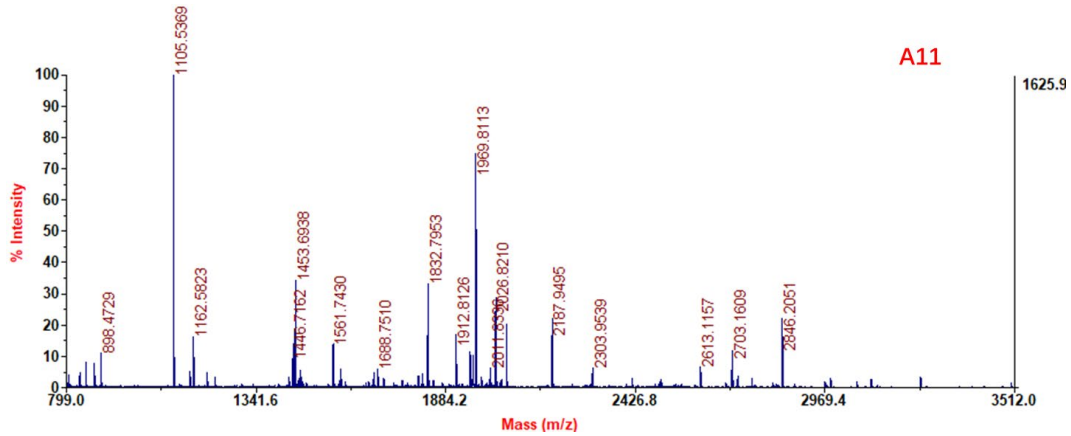

A6
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
3320.3
1162.5522
1463.6826
1562.7418
1572.7185
1174.5059
1093.5194
836.4060
898.4741
1393.6196
1319.6217
1711.7771
1816.7963
1888.8459
1975.8662
2056.9028
2133.0125
2229.9548
2689.1641
2777.2087
2879.3198
3100.3042

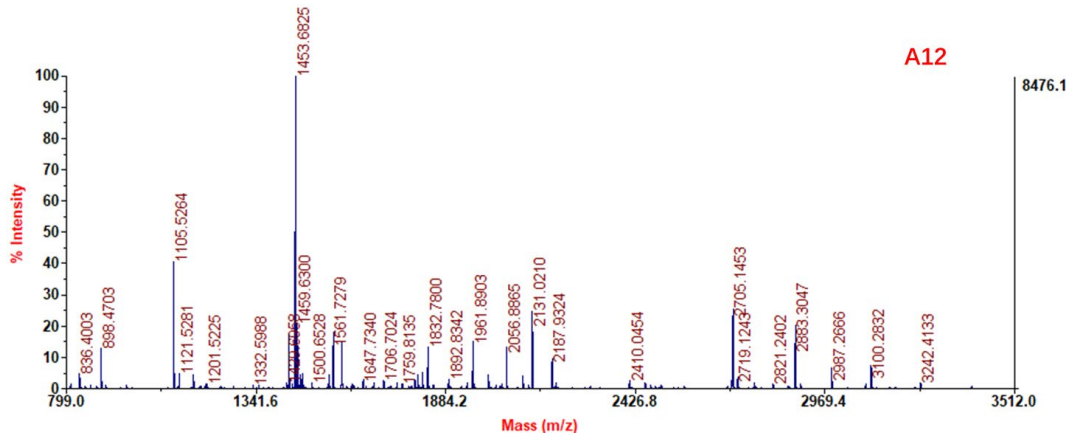

A7
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
3423.5
1453.6768
1105.5219
1121.5300
836.4004
898.4650
1561.7278
1706.7025
1832.7781
1961.8871
2056.8862
2131.0159
2187.9312
2705.1367
2883.2991
2988.2654
3100.2839

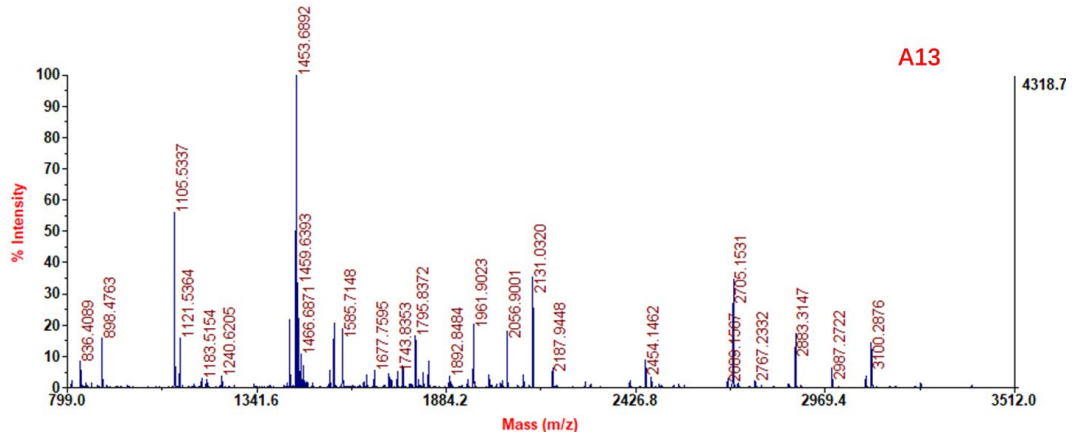

A8
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
5289.9
1453.7002
1105.5416
836.4141
898.4820
1121.5464
1240.6324
1459.6503
1466.7000
1586.7167
1677.7736
1743.8489
1795.8527
1804.8548
1892.8643
1961.9193
2056.9182
2131.0510
2187.9673
2455.1646
2705.1753
2883.3391
2987.2993
3100.3167

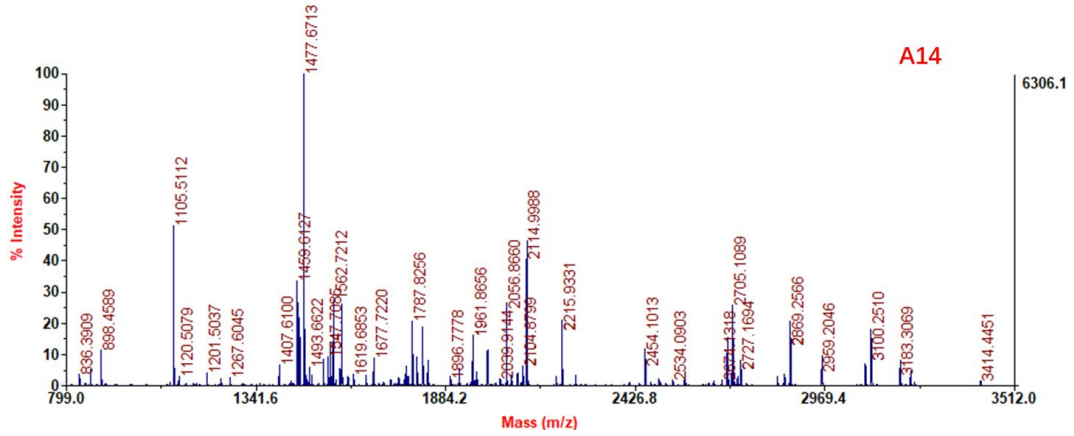

A9
% Intensity
Mass (m/z)
1105.5353
836.4081
898.4757
971.4168
1196.5969
1276.5941
1337.5671
1453.6936
1459.6421
1500.6667
1561.7423
1585.7169
1655.7433
1706.7178
1832.7998
1961.9050
2056.9026
2131.0396
2187.9473
2705.1646
2883.3228
2988.2859
3517.4

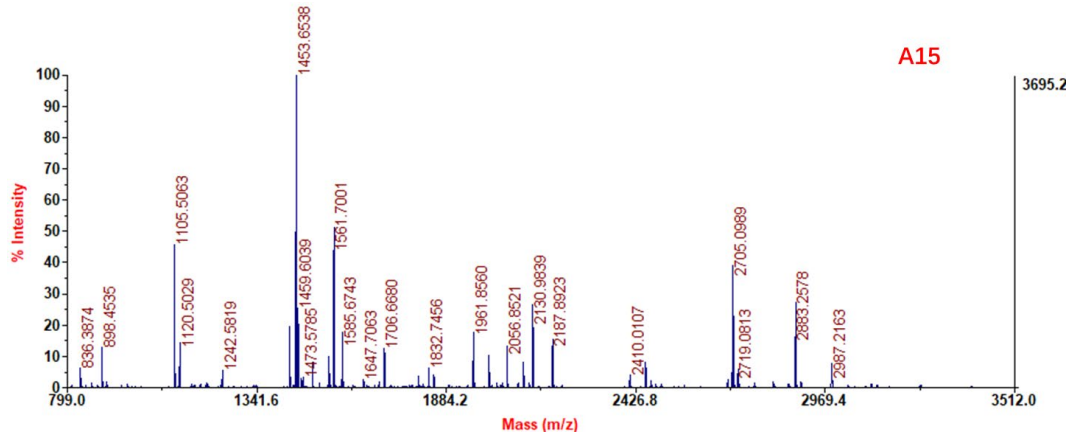

A10
% Intensity
Mass (m/z)
1453.6964
1105.5356
1162.5815
836.4084
898.4767
971.4191
1332.6149
1435.6449
1473.6230
1561.7454
1647.7544
1706.7192
1832.8021
1892.8579
1961.9137
2056.9111
2131.0459
2187.9580
2410.0762
2501.1021
2705.1753
2767.2490
2821.2747
2883.3401
2987.3013
3100.3162
5757.6

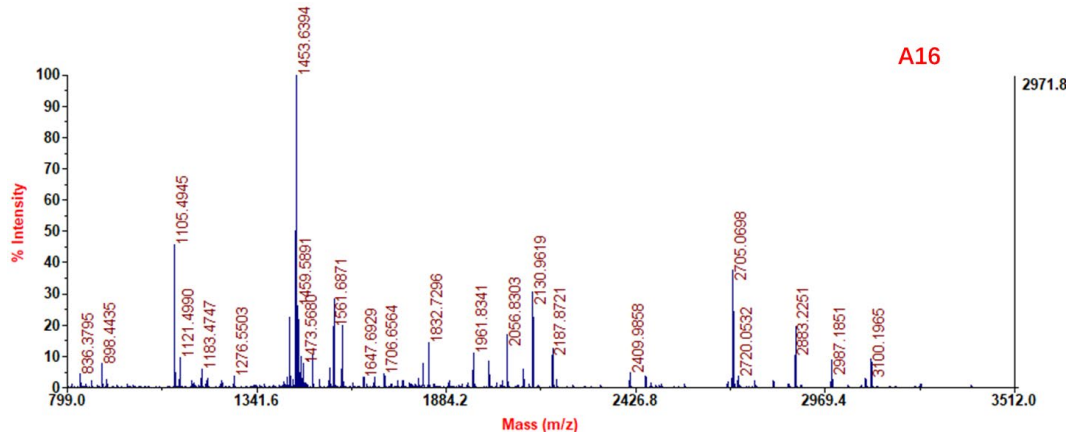

A11
% Intensity
Mass (m/z)
1105.5369
898.4729
1162.5823
1446.7162
1453.6938
1561.7430
1688.7510
1832.7953
1912.8126
1969.8113
2026.8210
2187.9495
2303.9539
2613.1157
2703.1609
2846.2051
1625.9


A12
% Intensity
Mass (m/z)
1453.6825
1105.5264
836.4003
898.4703
1121.5281
1201.5225
1332.5988
1459.6300
1500.6528
1561.7279
1647.7340
1706.7024
1759.8135
1832.7800
1892.8342
1961.8903
2056.8865
2131.0210
2187.9324
2410.0454
2705.1453
2719.1243
2821.2402
2883.3047
2987.2666
3100.2832
3242.4133
8476.1


A13
4318.7
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
1453.6892
1105.5337
1459.6393
1466.6871
1121.5364
1585.7148
1677.7595
1743.8353
1795.8372
1892.8484
1961.9023
2056.9001
2131.0320
2187.9448
2454.1462
2705.1531
2767.2332
2883.3147
2987.2722
3100.2876
836.4089
898.4763
1183.5154
1240.6205


A14
6306.1
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
1477.6713
1105.5112
1459.6127
1562.7212
1787.8256
1961.8656
2056.8660
2114.9988
2215.9331
2454.1013
2534.0903
2705.1089
2727.1694
2869.2566
2959.2046
3100.2510
3183.3069
3414.4451
836.3909
898.4589
1120.5079
1201.5037
1267.6045
1407.6100
1493.6622
1619.6853
1677.7220
1896.7778


A15
3695.2
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
1453.6538
1105.5063
1561.7001
1459.6039
1585.6743
1706.6680
1832.7456
1961.8560
2056.8521
2130.9839
2187.8923
2410.0107
2705.0989
2719.0813
2883.2578
2987.2163
836.3874
898.4535
1120.5029
1242.5819
1473.5785
1647.7063


A16
2971.8
% Intensity
Mass (m/z)
799.0
1341.6
1884.2
2426.8
2969.4
3512.0
1453.6394
1105.4945
1459.5891
1561.6871
1832.7296
1961.8341
2056.8303
2130.9619
2187.8721
2409.9858
2705.0698
2720.0532
2883.2251
2987.1851
3100.1965
836.3795
898.4435
1121.4990
1183.4747
1276.5503
1473.5680
1647.6929
1706.6564

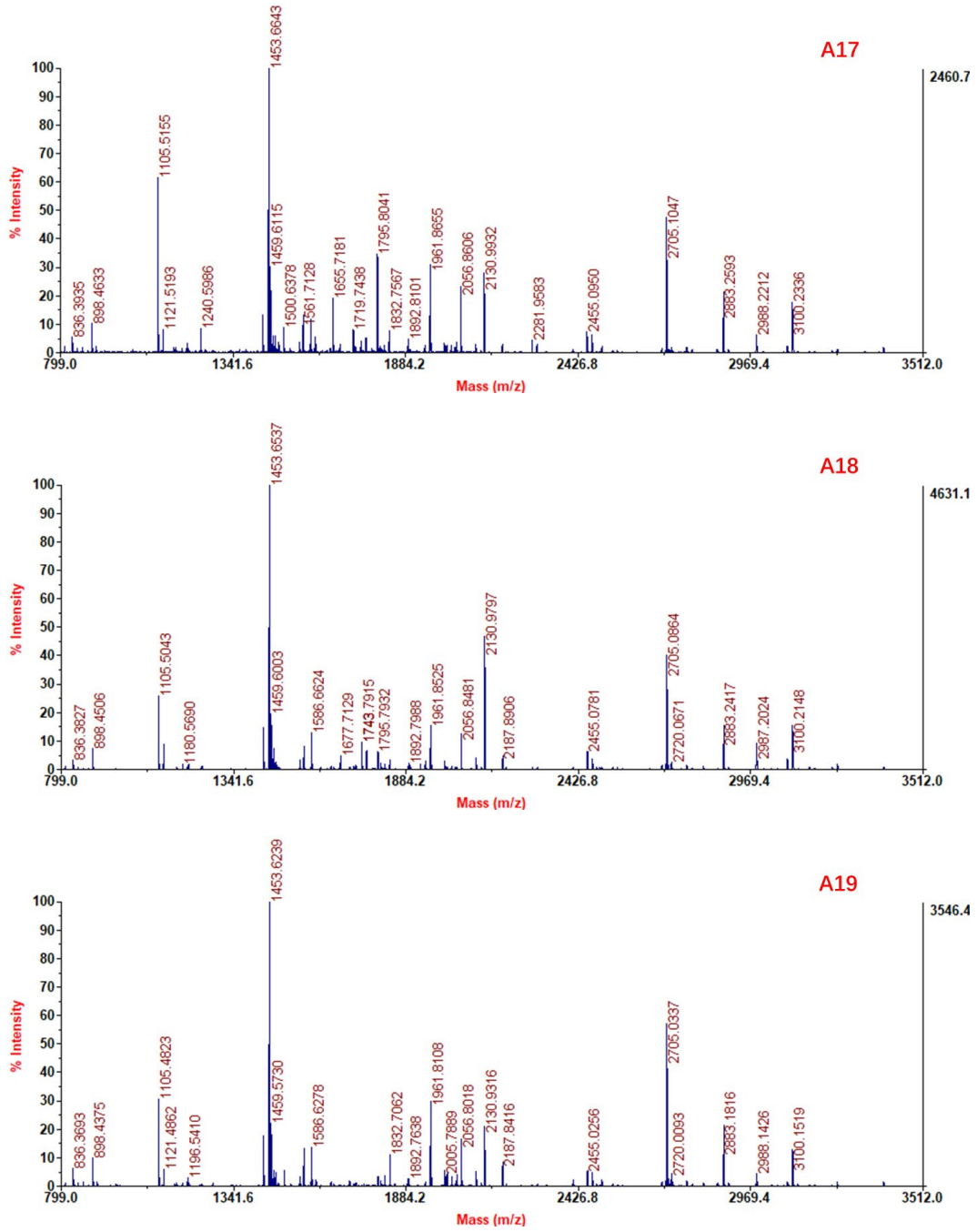


Fig. S1. MALDI-TOF spectra of the bone fragments.

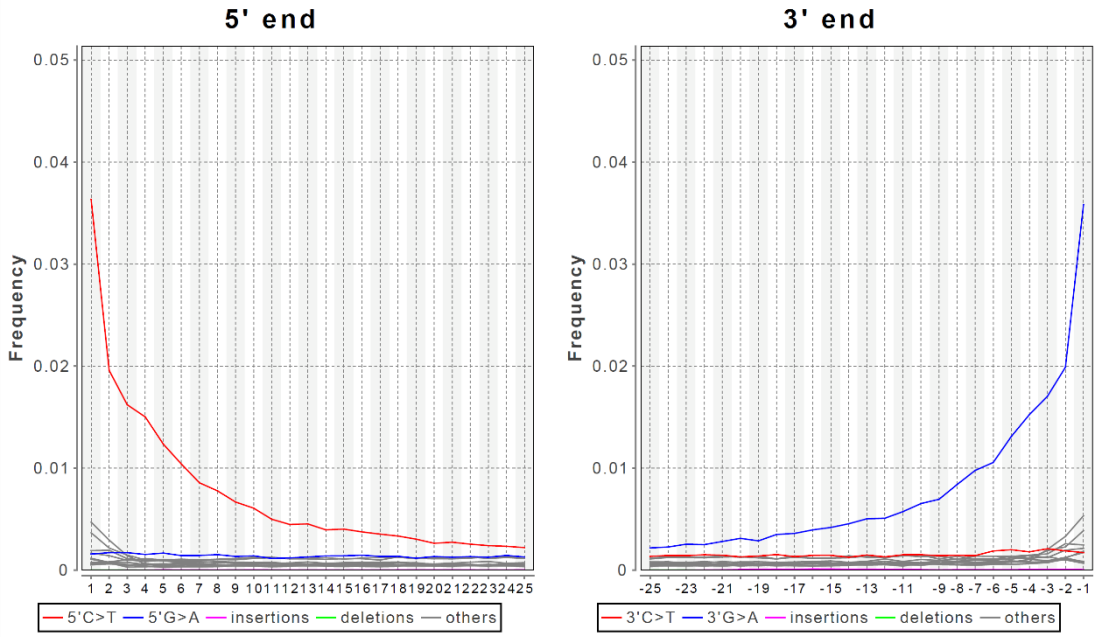


Fig. S2. DNA damage level of A14 measured by the rate of cytosine deamination-based misincorporation of bases as a function of position on reads. Red and blue lines represent 5'-end (C>T) and 3'-end (G>A) misincorporations, respectively.

Table S1. Taxonomic identification results of 19 samples by ZooMS.

Test No.	Sample No.	Subsample weight (mg)	Peptide mass fingerprint data (*m/z*)												ZooMS result
			COL1 α1 507-518 (P1)	COL1 α2 978-990 (A)	COL1 α2 978-990 (A')	COL1 α2 484-498 (B)	COL1 α2 502-519 (C)	COL1 α2 292-309 (P2)	COL1 α2 793-816 (D)	COL1 α2 454-483 (E)	COL1 α1 585-617 (F)	COL1 α1 585-617 (F')	COL1 α2 757-789 (G)	COL1 α2 757-789 (G')	
A1	2021SJ13	60	1105			1453			2131	2820	2883				*Sus scrofa*
A2	2021SJ06	26	1105			1453		1647	2131	2820	2883	2899			*Sus scrofa*
A3	2021SJ04	43	1105	1180		1453	1550	1647	2131		2883	2899			*Sus scrofa*
A4	2021SJ01	47	1105	1180		1453	1550		2131		2883	2899			*Sus scrofa*
A5	2021SJ08	69	1105			1453	1550	1647	2131		2883	2899		3033	*Sus scrofa*
A6	2021SJ02	36			1319	1463				2212	2252			3113	*Aves*
A7	2021SJ03	49	1105			1453	1550	1647	2131		2883	2899			*Sus scrofa*
A8	2021SJ07-1	39	1105			1453		1647	2131		2883	2899			*Sus scrofa*
A9	2021SJ05	42	1105	1180	1196	1453	1550		2131		2883				*Sus scrofa*
A10	2021SJ09	42	1105			1453	1550	1647	2131		2883	2899			*Sus scrofa*
A11	2021SJ10	58	1105			1453									Fail
A12	2021SJ11	37	1105			1453	1550	1647	2131		2883	2899		3033	*Sus scrofa*
A13	2021SJ12-1	46	1105			1453	1550	1647	2131		2883	2899			*Sus scrofa*
A14	2021SJ12-2	77	1105			1477	1580	1619	2115	2832	2869				*Homo sapiens*
A15	2021SJ12-3	45	1105			1453		1647	2131	2820	2883	2899			*Sus scrofa*
A16	2021SJ12-4	53	1105			1453		1647	2131	2820	2883	2899		3033	*Sus scrofa*
A17	2021SJ12-5	32	1105	1180		1453	1550	1647	2131		2883	2899			*Sus scrofa*
A18	2021SJ07-2	40	1105	1180		1453			2131	2820	2883	2899			*Sus scrofa*
A19	2021SJ12-6	43	1105	1180	1196	1453	1550		2131	2820	2883	2899			*Sus scrofa*

Table S2. Results of mtDNA HVR fragments extracted from bone remain of A14 and blood samples of the putative parents.

Sample	Mutation sites in mitochondrial HVR									
	16086	16136	16183	16184	16189	16223	16274	16309	16311	16362
rCRS	T	T	G	C	T	C	C	A	T	T
A14	T	T	G	T	T	T	C	A	C	C
pM	T	T	G	T	T	T	C	A	C	C
pF	C	C	C	C	C	-	T	G	T	C

Notes: The table lists the mutated sites of mitochondrial HVR-I sequences detected in the three specimens compared with Cambridge reference sequence (rCRS)[8]

Table S3. Results of Y chromosome haplogroup analysis of A14 and pF.

Haplogroups	Covered SNPs on the diagnostic positions of ISOGG database	
	A14	pF
O2a2b	M1690	F130, M1578, E303, M1585, P164, M1607, FGC16793, E488, FGC16799, CTS11109, M1729
O2a2b1	CTS11580	M1519.1, M1546, M9120, CTS7245, M1665, CTS8881, M1691, ACT533
O2a2b1a	n/a	n/a
O2a2b1a1	E284	M1516, E286, M1636, F14615, F649, F14692, CTS12991, F7810
O2a2b1a1a	M1545, CTS2810, Y9142, F363	F5, Z25787, M1521, F37, M1542, E273, M1545, F127, F139, M1567, M1591, M1597, CTS4497, M1622, CTS6104, M1647, F484, Y9139, Y9142, E530, M1704, CTS10672, CTS11637, CTS1377, F363
O2a2b1a1a5	M1532, M1726	CTS7316
O2a2b1a1a5b	n/a	F16367, A9457

References:

[1] Ning C, Wang CC, Gao S, et al. (2019) Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential Source of Indo-European Speakers in Iron Age Tianshan. Curr. Biol. 29: 2526-2532 e4 http://doi.org/10.1016/j.cub.2019.06.044.

[2] Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K (2016) EAGER: efficient ancient genome reconstruction. Genome Biol. 17: 60 http://doi.org/10.1186/s13059-016-0918-z.

[3] Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9: 88 http://doi.org/10.1186/s13104-016-1900-2.

[4] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-9 http://doi.org/10.1093/bioinformatics/btp352.

[5] Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012) Ancient admixture in human history. Genetics 192: 1065-93 http://doi.org/10.1534/genetics.112.145037.

[6] Kennett DJ, Plog S, R. J. George, et al. (2017) Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nat. Commun. 8: 14115 http://doi.org/10.1038/ncomms14115.

[7] Jeong C, Wikin S, Amgalantugs T, et al. (2018) Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. Proc. Natl. Acad. Sci. U. S. A. 115: E11248-E11255 http://doi.org/10.1073/pnas.1813608115.

[8] Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23: 147 http://doi.org/10.1038/13779.

Appendix 3 - Unpublished ZooMS results

As part of the investigation in this dissertation, brief descriptions of three pilot studies are provided below. The links to the raw data and results for Jiegedong and Xibaimaying can be found in Appendix 5.

Jiegedong Cave

Jiegedong Cave, located in the southern piedmont of the Qinling Mountains, was excavated in 2018 and 2019. The exposed area was 27 m^2, with an average deposit depth of 1.6 m, divided into 13 layers. Layers 1 and 2 are modern horizons; layers 3-5, ~30-15 ka; layers 6-8, ~50-70 ka; layers 9 and 10 are >100 ka. The lithic assemblage is dominated by a core-and-flake (small flakes) industry throughout the deposit. The excavation revealed interesting features such as a clear tread surface under layer 4, a knapping spot, and two fireplaces near the cave entrance. Additionally, two teeth of modern humans were discovered in layers 3-4. All sediments were water-sieved, resulting in the collection of over 10,000 small-size finds.

Recognising the potential of human remains in the unidentified bones collection, 326 fragmented bones from seven layers and nine units at Jiegedong were sampled for the ZooMS pilot study. Among the 326 analysed bones, 39 (12%) failed to yield sufficient collagen for identification. Jiegedong, being a cave site, exhibits a high level of collagen preservation. ZooMS-based determinations include 37% cervids (n=121) and 47% Bos/Bison and Bubalus (n=153), aligning with the preliminary morphological zooarchaeological examination (Zhang, personal communications, 2022). The morphological identification indicates a more diverse fauna composition at Jiegedong, except for Leporidae, which only appeared in ZooMS identifications. Cervids and bovids consistently dominated the Jiegedong fauna assemblage, and no significant differences were observed in deamidation levels among different taxa. The analysed samples span about 100 ka across nine layers. It is noteworthy that the ZooMS failure rate slightly increases with depth, suggesting collagen degradation with age.

Xibaimaying

Xibaimaying is an open-air site within the Nihewan Palaeolithic complex. Previous surveys categorised it as displaying a small flakes industry until the late Upper Palaeolithic, contrasting to the microlithic technique found at the nearby site, Youfang, in its late Upper Palaeolithic layers (24-14 ka BP). Two campaigns in 2015 and 2017 resulted in an excavated area of up to 100 m^2, recovering over 10,000 lithic and bone remains. New OSL dates indicate that the small-flakes-contained deposit at Xibaimaying is approximately 20 ka older than the microlithic-contained layer of Youfang, dating to 50-30 ka BP. Evidence such as cut markers on bones and numerous large lithic choppers suggests that Xibaimaying was used as a butchery camp by humans.

For the ZooMS pilot study, 39 bone specimens from layers 17-20, from the fauna assemblage of the 2015 campaign, were sampled. Of these, 25 yielded detectable collagen for ZooMS identification (64%), including horses, cervids, and bovids. Ongoing zooarchaeological research will provide a basis for comparison with ZooMS results. It is recognised that open-air sites may not preserve collagen as well as caves; however, the 64% identification rate at Xibaimaying, a loess deposit in the Nihewan, suggests the feasibility of additional ZooMS projects.

Bashan

Bashan, a recently excavated open-air site in Shandong Province, has undergone three consecutive years of fieldwork since its discovery in 2020, with a cumulative area of 225 m^2. Over 26,000 specimens, including more than 16,000 lithic and over 1,000 bone remains, were unearthed. The 8 m sequence is divided into 14 layers, dating to 100-60 ka BP (OSL data).

Sixty-three samples were selected for ZooMS analysis, mostly from layers 3-5, dating back to approximately 50 ka BP. Unfortunately, none of the 63 samples resulted in any positive ZooMS spectrum, indicating poor biomolecular preservation at Bashan.

Appendix 4 - Supplementary data

Supplementary files of Section 3.2

This dataset consists of Table S1 and other files relevant to Section 3.2.

Wang, Naihui (2024), “Supplementary data for the doctoral dissertation *Development and application of palaeoproteomics (ZooMS) on Palaeolithic assemblages from Europe and East Asia*”, Mendeley Data, V1, doi: 10.17632/fzyhj6vjzt.1

Jinsitai and Yumidong

All ZooMS spectra and result files have been uploaded in Mendeley data.

Wang, Naihui; Douka, Katerina (2024), “ZooMS data of Yumidong and Jinsitai caves”, Mendeley Data, V2, doi: 10.17632/ssf27rywhh.2

Vogelherd

All ZooMS spectra and result files have been uploaded in Mendeley data.

Wang, Naihui; Douka, Katerina (2024), “Vogelherd ZooMS data”, Mendeley Data, V2, doi: 10.17632/9jp4jdzy7k.2

Jiegedong

All ZooMS spectra and result files have been uploaded in Mendeley data.

Wang, Naihui (2024), “ZooMS data of Jiegedong Cave”, Mendeley Data, V2, doi: 10.17632/692pmcrytd.2

Xibaimaying

All ZooMS spectra and result files have been uploaded in Mendeley data.

Wang, Naihui (2024), “ZooMS data of Xibaimaying site”, Mendeley Data, V2, doi: 10.17632/c5ptd8vk7z.2