# Investigation of ancient proteins in archaeological material

**Dissertation**

in Partial Fulfilment of the Requirements for the Degree of

“Doctor of Philosophy” (PhD)

Submitted to the Council of the Faculty of Biological Sciences

of the Friedrich Schiller University Jena

by Ashley Scott, M.Sc. Biochemistry and Molecular Biology
born on 04 February 1985 in Gainesville, USA

Gutachter:

1. Prof. Dr. Christina Warinner, Friedrich Schiller University, Max Planck Institute for Evolutionary Anthropology
2. Prof. Dr. Johannes Krause, Friedrich Schiller University, Max Planck Institute for Evolutionary Anthropology
3. Prof. Dr. Matthew Collins, Cambridge University

*Beginn der Promotion:* July 14, 2017

*Dissertation eingereicht am:* October 28, 2021

*Tag der öffentlichen Verteidigung:* May 11, 2022

**Abbreviations** .......... **4**
**1. Introduction:** .......... **5**
1.1 Background: Dietary reconstruction .......... 7
1.2 Dental calculus .......... 18
1.3 Methods used in this thesis .......... 20
**2. Aims of the Thesis** .......... **28**
**3. Overview of manuscripts and author contributions** .......... **30**
3.1. Manuscript A .......... 30
3.2 Manuscript B .......... 32
3.3. Manuscript C .......... 34
3.4. Manuscript D .......... 36
**4. Manuscript A** .......... **38**
**5. Manuscript B** .......... **45**
**6. Manuscript C** .......... **62**
**7. Manuscript D** .......... **71**
**8. Discussion** .......... **83**
8.1 Dietary protein recovery and identification .......... 83
8.2 Patterns in protein preservation .......... 87
**9. Conclusion** .......... **101**
**10. References** .......... **103**
**11. Summary** .......... **117**
**12. Zusammenfassung** .......... **119**
**13. Declaration of Honour (Eigenständigkeitserklaerung)** .......... **121**
**14. Curriculum Vitae** .......... **122**
**15. Acknowledgements** .......... **122**
**16. Appendix** .......... **123**

# Abbreviations

| | |
|---|---|
| BLG | β-lactoglobulin |
| $AS_1$-CN | alpha-$S_1$- casein |
| A-LAC | α-lactalbumin |
| OVAL | Ovalbumin |
| BGLU | β-1,3- glucanase/glucosidase |
| TURM | Turmerin |
| BCG | β- conglycinin |
| SBP | Sucrose binding protein |
| LMW-GS | Low molecular weight glutenin subunit |
| AAI | α- amylase/trypsin inhibitor |
| CNs | caseins |
| BLG II | β-lactoglobulin II |
| $AS_2$-CN | alpha-$S_2$- casein |
| B-CN | beta-casein |
| K-CN | kappa-casein |
| SA | serum albumin |
| LTF | lactoferrin/ lactotransferrin |
| LYZC | lysozyme C |
| $\delta^{13}C$ values | $^{13}C$ to $^{12}C$ ratio |
| $\delta^{15}N$ values | $^{15}N$ to $^{14}N$ ratio |
| Δ13C | the difference in $\delta^{13}C$ values of $C_{16}$ and $C_{18}$ |
| LP | lactase persistence |
| LNP | lactase non-persistence |
| GC-MS | gas chromatography-mass spectrometry |
| GC-C-IRMS | gas chromatography-combustion-isotope ratio mass spectrometry |
| MALDI-TOF MS | matrix-assisted laser desorption ionization time-of-flight mass spectrometry |
| LC-MS | liquid chromatography-mass spectrometry |
| PTMs | post-translational modifications |
| PSMs | peptide spectral matches |
| m/z | mass/charge |
| FASP | filter-aided sample preparation |
| EDTA | ethylenediaminetetraacetic acid |
| UA | urea |
| SDS | sodium dodecyl sulfate |
| DTT | dithiothreitol |
| IAA | iodoacetamide |
| NaCl | sodium chloride |
| TFA | trifluoroacetic acid |
| TEAB | triethylammonium bicarbonate |
| SPE | solid-phase extraction |
| ACN | acetonitrile |
| FA | formic acid |
| FDR | false discovery rate |
| AFs | amyloid fibrils |
| IgE | immunoglobulin E |
| CTL | c-type lectin |

# 1. Introduction:

The past decades have witnessed the rapid expansion of studies into ancient human diet using a range of diverse methods. Technical advances in the fields of ancient DNA, microscopy, isotopic chemistry, organic residue, and protein chemistry have allowed researchers to gain new insight into the economy and subsistence of ancient humans (Roffet-Salque et al. 2017; Hendy 2021; Katzenberg and Grauer 2018; Henry 2020; Mann et al. 2020). Integrating these methodologies with more traditional archaeological research and historical texts and iconography can provide insights into the human behavior and cultural choices that have allowed humans to adapt and modify their subsistence strategies in the past. Food consumption and nutrition patterns can help track the use, spread, and evolution of domesticated plant and animal species and food processing technology through time.

Conventionally, past human diets have been primarily studied through comparative morphological analysis of botanical and faunal assemblages. However, the recent application of molecular methods to the recovery and identification of ancient organic molecules has led to many discovery-based findings that have demonstrated the potential and limitations of these new methods when applied to the archaeological record (Zischler et al. 1995; Collins et al. 1995; Downs and Lowenstein 1995; McGovern et al. 1995; Fiedel 1996; Austin et al. 1997; Dongoske et al. 2000). Researchers have applied molecular methods to many artifact types and remains in an attempt to glean dietary information from the past. Bone, paleofeces, pottery, stone tools, and dental calculus are among the substrates studied to date (Katzenberg and Grauer 2018; Warinner et al. 2014; Maixner et al. 2018; Dunne et al. 2012; Hendy et al. 2018; Martisius et al. 2020; Korzow Richter et al. 2020).

The research presented in this thesis was produced by proteomic analysis of ancient dental calculus in conjunction with data from other archaeological methods, including genomic and microremain analyses. Dental calculus is mineralized plaque that adheres and accumulates on the surfaces of teeth, trapping biomolecules and food remains, which provides researchers with a direct line of evidence for the consumption of dietary items. This thesis will focus on ancient dietary proteins and the strengths and limitations inherent in paleoproteomic approaches to dietary studies. When considering the reconstruction of paleodiet, all methods employed, both new and traditional, have limitations. None of the methods discussed in this thesis can provide a complete picture of ancient diet, and it would be inappropriate to attempt a full dietary reconstruction using any or even all methods at our disposal. However, proteomics methodologies can offer several important advantages over other methods. Researchers have demonstrated the exceptional preservation potential of ancient proteins, particularly when compared to other biomolecules, and proteins have been identified from material from millions of years in the past (Rybczynski et al. 2013; Demarchi et al. 2016; Cappellini et al. 2019). Proteins are often able to offer a level of taxonomic and tissue specificity unavailable to other methods. Tandem mass spectrometry technology, in particular, is highly sensitive and can offer exceptionally accurate measurements even in complex mixtures of unknown proteins. Notably, the method also provides probabilistic matches for peptide matches, which allows for more statistical precision for protein and taxonomic identifications, an advantage not available for many archaeological methods.

A considerable hurdle to the paleoproteomic study of diet is that many mechanisms of protein preservation are still poorly understood. Some protein characteristics have been hypothesized to play a role in long-term preservation in the archaeological record, including protein structure or function, calcium-binding ability (Collins et al. 2002; Warinner et al. 2014a; Demarchi et al. 2016; Hendy et al. 2018b), and the presence of specific post-translational

modifications such as glycosylation (Ozcan et al. 2014). The most frequently recovered proteins are fibrous, insoluble proteins such as keratin, collagen, and elastins. In particular, mineralized fibrous proteins are often exceptionally well-preserved (Demarchi 2020). Collagen, the most abundant protein in animals, possesses a particularly stabilizing triple helical structure, and many collagen types form fibrillar macromolecular structures (Shoulders and Raines 2009). As expected, collagen proteins are the most recovered protein in paleoproteomic studies. In contrast, diet-specific protein recovery from archaeological material has been rare, and some early identifications do not hold up to more recent bioinformatic standards, and many reported proteins cannot be verified through independent data analysis.

Prior to the onset of this thesis research, unequivocal dietary remains from dental calculus primarily consisted of dairy proteins, namely β-lactoglobulin (BLG) (Warinner et al. 2014b, a). BLG continues to dominate dietary proteomic identifications (Yang et al. 2014; Jeong et al. 2018; Charlton et al. 2019; Wilkin et al. 2020; Bleasdale et al. 2021). However, the small number of unequivocal dietary proteins that have been identified compared to the large number of unidentified MS/MS scans that dominate paleoproteomic datasets had offered limited insight into the factors contributing to dietary preservation. Compared to commonly recovered proteins in archaeological bone and shell, BLG shares few similarities beyond its calcium-binding capacity.

Although the recovery of non-dairy dietary proteins has been rare thus far, this thesis aims to provide a better understanding of the factors that may influence the preservation and recovery of proteins and to identify potential biases introduced through the differential preservation of proteins. Through Manuscripts A, B, C, and D, this thesis expands the number of robustly identified ancient dietary proteins in dental calculus, ultimately resulting in the recovery of 20 different dietary proteins that could be taxonomically identified to the genus or species level from 10 different plant and animal sources (see Table 1), including milk, egg, wheat, sesame, soybean, turmeric, and banana. Many of these proteins had not been previously identified using paleoproteomics techniques.

In order to achieve greater protein recovery from dental calculus, this thesis included the development of a series of methodological improvements and optimizations to the protein extraction protocol. These efforts focused on reducing loss, preventing off-target chemical modifications, and improving protein purity. The resulting protocol has been made publicly available on the public repository Protocols.io: dx.doi.org/10.17504/protocols.io.7vwhn7e. Additional efforts were also made to improve the database completeness of milk proteins for common dairy livestock. These activities, aimed at improving the recovery and identifiability of ancient dietary proteins, are described in section 1.3.

Finally, this thesis examines patterns of dietary protein recovery and identification in ancient samples to date. The Discussion section of this thesis (section 8) presents a series of observations and hypotheses regarding the potential mechanisms of protein preservation in archaeological contexts. Having a deeper understanding of the types of dietary proteins that can be expected to preserve over time is a necessary next step for the field of paleoproteomics, particularly in the context of dietary identification from dental calculus. Until we understand the potential biases in our method, any attempt at dietary reconstruction through proteomic studies will be limited. While not all potential biases are relevant or important in every paleodietary study, those introduced through our analytical methods will affect the types of research questions we can ask or expect to answer. As a result, a deeper

understanding of the factors that influence protein survival and recovery will enable more robust studies focusing on relevant research questions, the appropriate use of scarce archaeological material, and new means of authentication.

## 1.1 Background: Dietary reconstruction

In the following sections, the main conventional archaeological methods that have been used to reconstruct aspects of ancient diets will be discussed, including archaeobotany, zooarchaeology, stable isotope analysis, and organic residue analysis. The emerging field of paleoproteomics is then presented, and the strengths and weaknesses of this new molecular technique are described.

Although multiple methods exist to reconstruct aspects of ancient diets, each of the methods available can only provide specific, often limited, information about the overall diet, yielding a snapshot into the diet of individuals, or a set of individuals, in a fixed time and place. Each method employed can suffer from the limitations inherent in any archaeological research; namely, assemblages are non-renewable and often involve low sample sizes and small starting amounts that can restrict any attempt at statistical analysis and lead to biases. It is also important to consider that all archaeological assemblages are biased from the start (Vanderwarker and Peres 2010). Biases are introduced from the initial formation of an

**Table 1. Dietary protein recovery from four manuscripts discussed in this thesis.** A total of 2,052 dietary-specific peptide spectral matches (PSMs) were identified from archaeological dental calculus. Milk proteins were the most frequently recovered proteins, and the majority ( 1756/1898) of these were recovered from individuals that participated in a pastoralist subsistence economy

| Food category | Proteins | | Taxonomic ID | PSMs |
|---|---|---|---|---|
| **Milk** | *Caseins* | $\alpha S_1$-casein ($AS_1$-CN) | Caprinae | 9 |
| | | $\alpha S_1$-casein ($AS_1$-CN) | Bovinae | |
| | *β-lactoglobulins* | β-lactoglobulin (BLG) | *Capra* | 1879 |
| | | β-lactoglobulin (BLG) | *Ovis* | |
| | | β-lactoglobulin (BLG) | Bovinae | |
| | | β-lactoglobulin I (BLGII) | *Equus* | 9 |
| | | α-Lactalbumin (A-LAC) | Caprinae | 10 |
| **Egg** | | Ovalbumin (OVAL) | *Gallus* | 17 |
| **Fruit** | | β-1,3- glucanase/glucosidase (BGLU) | *Musa* | 2 |
| **Tubers** | | Turmerin | *Curcuma* | 4 |
| **Seeds** | *11S globulin seed storage proteins* | 11S globulin isoform 2 | *Sesamum* | 29 |
| | | 11S globulin isoform 3 | *Sesamum* | 23 |
| | | 11S globulin isoform 4 | *Sesamum* | 16 |
| | | 11S globulin: Glycinin G1/G2 | *Glycine* | 12 |
| | | 11S globulin: Glycinin G4 | *Glycine* | 6 |
| | | 7S globulin: β- conglycinin (BCG) | *Glycine* | 10 |
| | | Sucrose binding protein (SBP) | *Glycine* | 2 |
| | | 2S albumin | *Sesamum* | 8 |
| **Grains** | | Low molecular weight glutenin subunit (LMW-GS) | *Triticum* | 13 |
| | | α- amylase/trypsin inhibitor (AAI) | *Triticum* | 3 |

archaeological site as behavior, and cultural practices dictate which dietary taxa will and will not be left behind. Decisions about butchering location, cooking, discarding, or storing dietary items will affect which items remain. Differential preservation and taphonomy will introduce further biases as weathering, soil pH, moisture, scavenging, plant intrusion, redeposition, trampling, and human activity can prevent recovery of archaeological items (Lee Lyman and Lyman 1994). Archaeological excavations and field recovery procedures will bias dietary studies further, as excavators must often decide what can be reasonably recovered. Researchers can quickly introduce considerable biases by choosing which samples to study, which analytical methods to employ, and how they interpret the resulting data.

### *1.1.1 Archaeobotanical studies of past diet*

Researchers studying ancient diets have often utilized recovered plant remains to investigate subsistence strategies, plant use and domestication, and the development of agriculture. Generally, researchers classify plant remains as either macroremains or microremains. Macroremains are visible with the human eye or with low-powered microscopy, while microremains require a high-powered microscope for identification (Wright 2010).

#### *1.1.1.a. Macroremains*

Macrobotanical remains have conventionally been one of the primary sources of data for the study of past diets. In general, there is a strong bias against the recovery of macrobotanicals, and many regularly consumed vegetables and fruits are highly perishable. Of the macroremains that preserve, there is a strong bias towards the recovery of wood, grains, and seeds, especially those that have carbonized. Archaeologists have also recovered more delicate macroremains such as tubers, stems, leaves, and fruits from archaeological sites with optimal environmental conditions for preservation. These conditions are rare and consist of environments with little to no oxygen or moisture, preventing bacterial decomposition of organic material, such as permanently anoxic waterlogged conditions such as bogs (Hillman 1986) and shipwrecks (Haldane 1993). There is also an increased chance of preservation of plant remains in very arid environments such as deserts. While exceedingly rare, these conditions have produced a wide variety of plant remains, including more delicate and highly perishable types of remains, such as desiccated flowers, fruits, and leaves. Likewise, the process of carbonization occurs through a low-oxygen environment. Charred plant remains are converted into charcoal (carbon), which prevents the natural decay process.

Archaeologists typically recover macroremains from archaeological contexts through hand collection, screening soil with fine mesh, or flotation. Flotation is the most common recovery method used for macrobotanical remains. This process uses either water or chemicals to free plant remains from their geological substrate. While sediments sink, less dense organic and charred materials float to the top, allowing separate recovery. Some more dense plant remains will sometimes sink into the sediment fraction, so both fractions are often dried and sorted for remains (Wright 2010).

#### *1.1.1.b. Microremains*

Microbotanical remains are plant remains that require high-power microscopy for identification and include pollen, starch granules, and phytoliths. Pollen is produced in the male reproductive organs of spermatophytes. Plants produce pollen with a wide diversity of morphologies, and these differences assist in taxonomic classification. In addition to pollen, most plants also form starch granules, which store chemical energy produced through photosynthesis within chloroplasts in leaves and stems and within amyloplasts in roots, tubers, rhizomes, and seeds. The morphology of starch granules can allow for identification,

although the starch granules found within chloroplasts are often non-diagnostic and therefore of limited use (Henry 2020a). The starch granules produced within amyloplasts are often larger and store carbohydrates as an energy source for the plant. In some cases, these granules have a sufficiently distinctive morphology to allow identification, but these species-level taxonomic identifications are generally rare (Mercader et al. 2018). Phytoliths are produced by some higher plants by absorbing silica from groundwater, which precipitates within the plant and solidifies into recognizable shapes and sizes that help identify plants taxonomically. These phytolith fossils survive long after the plant has decayed, which can be particularly advantageous. Fossilized silica can survive in environments that are not conducive to preserving macrobotanical structures, and phytoliths can give us a glimpse of dietary specimens that might otherwise remain invisible.

Flotation is the primary recovery technique for microremains. Due to the microscopic size of the remains, archaeologists typically take samples from specific locations within an archaeological site rather than process large amounts of soil. In addition to dry sediments and lake cores, microremains are recoverable from other contexts, including structural surfaces, ceramic vessels, dental calculus, paleofeces, and other archaeological artifacts (Hunt et al. 2001; Wilkinson et al. 2003; Piperno 2006; Tromp and Dudgeon 2015; Henry 2020b).

*1.1.1.c. Limitations of archaeobotanical studies*

While some plants produce characteristic microremains, most have a lower taxonomic value due to the significant overlap in morphology between and within plant species. Pollen is often only identifiable to the genus level (Day 2013). Starch and phytoliths can also be limited in their taxonomic specificity. Phytoliths have the lowest taxonomic value, but they can be used for broader classifications such as grasses or cereals, providing insight into dietary habits.

Macrobotanical remains are strongly biased towards the recovery of carbonized grains and seeds. In general, macrobotanical remains rarely preserve, and many regularly consumed structures are highly perishable. While carbonization can protect structures from natural decay, mechanical damage from food processing or environmental conditions can also complicate the identification of taxa (Vanderwarker and Peres 2010). The carbonized remains that do preserve are often found near hearths. The types of foods that are likely to be charred require processing before consumption, such as cereals. Since this preserve more readily, they are more likely to be overrepresented. Plants with fragile structures or plants eaten raw or away from archaeological features such as hearths are unlikely to be recovered or sampled, and they will be under-represented or invisible in the plant assemblages.

Differential preservation of microremains can also introduce bias. Heat can cause damage or loss of features used to identify plant species. Pollen only survives in an uncharred form. Processes such as heat, water absorption, and desiccation can change the morphological characteristics of starch granules, complicating identification. Starch granules themselves preserve differentially by both size (Haslam et al. 2009) and species through the preferential breakdown of some plants by bacteria (Hutschenreuther et al. 2017). Silica does not accumulate the same in all plants. Some plants, like grasses, produce an abundance of phytoliths, while others produce few, if any, and can remain invisible. Some plants produce more phytoliths with high taxonomic value, such as Poaceae. Additionally, phytolith morphotypes also preserve differentially. Phytoliths start to dissolve soil pH above 8. Morphotypes dissolve at different rates, influenced by their shape, with more compact forms dissolving less than those with a higher surface/volume ratio (Cabanes and Shahack-Gross 2015). These factors can introduce biases that can complicate any interpretation of the plant

assemblages and their relation to ancient diets.

Even if conditions are ideal for preservation, sampling methods can introduce a bias in the recovery of botanical remains. Recovery will vary with technique, mesh size, and the type, size, and fragility of remains (Wright 2010). Flotation techniques are too labor-intensive to be practical for large-scale soil sampling. Instead, smaller subsamples are often selected, and consequently, the recovered plant assemblages will not fully represent everything deposited. While phytoliths and starch granules are more likely to remain in their deposited location, wind, water, or soil erosion can easily carry pollen from its original location (Evans et al. 1999), limiting interpretation of human involvement. Piperno demonstrated the differential recovery of microremain types from various substrates such as soil, paleofeces, dental remains, and artifacts (Piperno 2006).

The identification of botanical remains is an inherently comparative method. Analysts visually compare archaeological remains to known, often modern, specimens. Extensive and reliable comparative reference collections are indispensable for accurate identification. Due to the comparative nature of most botanical identifications, there is always some inherent subjectivity involved in taxonomic designations.

### *1.1.2 Zooarchaeological studies of past diet*

Zooarchaeological studies of animal remains have conventionally played an essential role in the investigation of ancient human diets. Data from these studies can provide insight into ancient subsistence strategies, what role specific animal species had in these economies, and the technological adaptations involved in exploiting animal resources.

Numerous taphonomic processes can hinder or aid the preservation of skeletal remains. These processes include cultural transformations caused by human processing such as butchery and cooking and abiotic factors including soil pH or exposure to sunlight and moisture. Post deposition disturbances such as scavenging, erosion, intrusive burials, and plant intrusion can also lead to the destruction of archaeological remains. Even in ideal conditions, some skeletal elements tend to survive better taphonomically, such as those that are more calcified (e.g., teeth, antlers, bones, or shells) (Lee Lyman and Lyman 1994; Peres 2010). Larger, more dense bones are also more likely to survive well, although they are more prone to fragmentation. These factors can bias faunal assemblages towards preserving and recovering larger animal bones and inflate count values, thereby inflating their relative representation (Lee Lyman and Lyman 1994).

Archaeologists typically recover faunal remains through dry screening of excavated soils, and mesh size will determine the size of recovered specimens. Mesh size can easily bias the recovery of smaller elements or fragments (Peres 2010). Like archaeobotanical studies, the traditional and most common method of faunal species identification is through visual inspection and designation based on taxonomically diagnostic morphological features. These identifications are made through comparison with a reference collection of faunal remains.

#### *1.1.2.a. Zooarchaeological studies of secondary products*

Zooarchaeological studies of subsistence typically focus on the relative abundance of specific animal species within an assemblage. If present, researchers can use sex-specific and age-specific features to infer herding strategies or the use of secondary products such as milk (Vigne and Helmer 2007). For example, optimal herd managing strategies include the culling of animals at specific life stages. Animal herds used primarily for meat consumption often

show culling patterns with an increased frequency of culled young and sub-adult males, typically between one and two and a half years (Marciniak 2011). Typically for goats and sheep used primarily for milk, there will be a high kill-off of the very young (often under two months of age), and herds will be composed of mostly older females (Vigne and Helmer 2007). However, accurately assessing these kill-off patterns requires a well-preserved faunal assemblage of an appropriate statistical size (Albarella et al. 2017). Both sample size and preservation quality can influence the validity of inferences made about animal exploitation strategies concerning ancient human diet.

1.1.2.b *Limitations of zooarchaeological studies*
As with any set of archaeological remains, zooarchaeological assemblages are inherently biased and unlikely ever entirely to reflect the past use of animals for human subsistence. Soft tissue rarely survives, and faunal assemblages typically consist of the most durable skeletal features. Remains recovered from settlements are often food-processing by-products, and butchering, transport, and discard patterns will affect the availability of some skeletal elements.

In ancient populations that are wholly or even seasonally mobile, zooarchaeological data can be challenging to interpret. These mobility patterns can obscure culling patterns, and statistically robust conclusions about herd management strategies or composition are impossible without a larger sample size than is often available with fully mobile populations that lacked permanent settlements (Gamble 1978). In these cases, assumptions about herd culling patterns or composition would be based on faunal remains found only within human burials. These grave burials reflect what the society chose to be buried with, not necessarily the full spectrum of exploited animals or how ancient societies utilized them (Russell 2011).
Zooarchaeologists face many challenges when quantifying a faunal assemblage. Differential preservation, fragmentation, and identifiability of skeletal elements and species can easily affect an assemblage's quantification (and subsequent statistical analysis), and researchers often use different methods to calculate these values (Peres 2010; Lambacher et al. 2016). The appropriateness of each measure is relative to the research question, and debate about the strengths and limitations of each measure is still ongoing (Peres 2010; Domínguez-Rodrigo 2012). There is also the concern about the validity of faunal identification and the reproducibility of zooarchaeological studies due to a lack of broadly accepted identification standards (Wolverton 2013). Like archaeobotanical studies, zooarchaeological studies are comparative, and therefore, to some degree, always subjective. Issues of intra- and inter-species variation in morphology can complicate or reduce identifications. Accurate identifications often depend on the skill and experience of the researcher as well as the size and quality of the faunal reference collection.

### *1.1.3 Stable isotopic studies of past diet*
Stable isotope studies have been a standard tool for archaeologists studying ancient diets since the first studies in the 1970s (Vogel and van der Merwe 1977; van der Merwe and Vogel 1978). The ratio. of stable isotopes for nitrogen and carbon in a food source undergo predictable patterns of fractionation as it incorporates into the consumer's skeleton, which can help differentiate the types of, and relative importance of the isotopically distinct foods consumed by an individual. Mass spectrometers are used to separate a sample by its elemental isotopes through their mass-to-charge ratio, providing researchers with the ratio of these isotopes relative to each other. Archaeological researchers have applied bulk and compound-specific stable isotope analyses to various dietary reconstruction studies (Katzenberg and Grauer 2018).

*1.1.3.a Stable isotopic substrates*

Various body tissues are better suited for specific research questions. Bone collagen is the most common tissue tested in investigating ancient human diets through stable carbon and nitrogen ratios. Collagen's stability and long-term survival potential are well documented (Lee-Thorp and Sponheimer 2003; Collins et al. 2002; Dobberstein et al. 2009). Bone collagen is remodeled over time with amino acids derived from dietary sources, providing researchers with a snapshot into protein consumed by the individual during the last few years before death. In comparison, collagen from tooth dentine does not remodel after formation in childhood. The isotopic composition is preserved and can be recovered to generate data about an individual's diet during tissue formation (Eriksson 2013). Combining different tissue types is a useful way to show potential isotopic changes within an individual's lifetime. In addition to collagen, carbonate ($CO_3$) found within biogenic apatite can also be used to source isotopic carbon ratios and is found in bone, tooth enamel, and dentine. The carbonate found within these sources is extraordinarily stable (Sponheimer and Lee-Thorp 1999; Lee-Thorp and Sponheimer 2003) and can be helpful in contexts where collagen is too degraded for analysis. Additionally, while the carbon in collagen is derived chiefly from dietary proteins, the carbonate carbon is derived from dissolved bicarbonate in the blood, sourced from dietary carbohydrates, lipids, and protein (Katzenberg and Grauer 2018). This can provide a better perspective into an individual's entire diet, including low protein foods (Krueger and Sullivan 1984; Ambrose and Norr 1993; Tieszen and Fagre 1993).

*1.1.3.b. Relative isotopic values and diet*

Stable isotopes are often used to distinguish $C_3$ plants from $C_4$ plants; their differing photosynthetic pathways alter their carbon fixation pathways, which affects their $^{13}C$ to $^{12}C$ ratio ($\delta^{13}C$ value). Plants that have evolved in different climatic regions have different $\delta^{13}C$ values. $C_4$ plants evolved in hot and dry environments and consist mostly of grasses, and their photosynthetic pathways are more efficient, reducing water loss compared to$C_3$ plants. Although $C_3$ plants make up the vast majority of plants on earth, a few economically important $C_4$ dietary species, such as maize, sorghum, and millet, have been domesticated by humans, and increasing $\delta^{13}C$ values in human bones can reflect an increased reliance on these dietary taxa (Vogel and van der Merwe 1977). The ratio of $^{15}N$ to $^{14}N$ (the $\delta^{15}N$ value) isotopes corresponds broadly to the consumption of dietary sources at various trophic levels. Plants exist at the lowest trophic levels, while carnivores exist at the top. Generally, individuals with a high $\delta^{15}N$ value are likely to have consumed more dietary resources higher in the food chain (i.e., meat) than individuals with a lower value. The proportion of meat to plant resources can be inferred by these isotopic values.

These studies can provide broad generalizations about ancient diets based on consumed protein, but they provide low resolution about the specific foods consumed. Additionally, although bulk $\delta^{15}N$ values can show if an individual's diet was high in animal-derived protein, this method cannot reliably distinguish between the consumption of milk or meat from domesticated animals (Sealy 2001). Researchers have used compound-specific analysis of $C_{16:0}$ and $C_{18:0}$ fatty acids using GC/C/IRMS (gas chromatography combustion isotope ratio mass spectrometry) to distinguish ruminant dairy from ruminant meat based on the relative $\delta^{13}C$ value ratios of fatty acid markers (Craig et al. 2012; Salque et al. 2013; Carrer et al. 2016; Grillo et al. 2020), discussed below in section 1.1.5.

Researchers may also use stable isotope analysis to distinguish between diets based on different ecosystems. Ecosystems have distinct isotopic values, and therefore the environmental origins may be inferred from these values. For example, bulk collagen $\delta^{13}C$

and $\delta^{15}N$ can be used to distinguish marine and terrestrial dietary sources (Lovell et al. 1986), while compound-specific stable acid analysis can provide more precise estimates for relative dietary contributions from marine and terrestrial sources (Corr et al. 2005). Dissolved carbonate is the primary carbon source for marine organisms, and they typically have higher $\delta^{13}C$ values than terrestrial $C_3$ organisms that use atmospheric $CO_2$ as their primary carbon source. The $\delta^{13}C$ value of terrestrial $C_4$ organisms is higher than both terrestrial $C_3$ and marine organisms (Eriksson 2013). $\delta^{13}C$ values also vary with salinity, and freshwater environments typically have similar values as those from $C_3$ terrestrial environments. However, aquatic carbon and nitrogen cycles are complex and context-dependent, and isotopic values can vary substantially (Eriksson 2013), complicating data interpretation.

*1.1.3.c. Limitations of stable isotope studies*

Like other analyses discussed in this thesis, stable isotope values are relative rather than absolute and are open to interpretation. For reliable results, the isotopic values of the dietary items studied must be well-characterized, and isotopic baselines for the studied environment should be considered, both geographically and temporally (Fernandes and Jaouen 2017). Unfortunately, this can be challenging for archaeological studies because acquiring a statistically appropriate sample set to establish these baselines is not always possible. Additionally, bulk isotope data can be limited in its ability to discern dietary patterns in human populations from isotopically complex landscapes due to isotopic equifinality, a scenario in which different combinations of dietary isotopic values can result in equivalent bulk $\delta^{13}C$ values and an inability to distinguish dietary components unequivocally (Bogaard and Outram 2013). Bayesian mixing model software has been utilized to estimate different contributions of dietary components through probability distributions (Fernandes et al. 2014), but these approaches are highly reliant on the overall dietary ecology being precisely defined and well understood.

Although clear isotopic signatures based on photosynthetic pathways, sources of water, and nitrogen use have been identified and accumulated into databases, there are often mismatches between expected theoretical and empirically observed values. Various complex factors can affect isotopic signatures, and therefore, interpretation of results can often be challenging (Marshall et al. 2007). Environmental factors such as aridity, salinity, and water stress, and anthropogenic factors such as cooking, heating, the use of manure as a fertilizer, nutritional stress, and disease can all affect stable isotopes (Ambrose 1991; Fogel et al. 1997; Bogaard and Outram 2013; Eriksson 2013) and lead to incorrect interpretations. The larger archaeological context is vital, and without some *a priori* knowledge about plant and animal species available for consumption at a site, accurate interpretation of isotopic data and quantitative isotopic studies may be impossible (Mahajan and Sathe 2020).

### *1.1.4 Paleogenetic studies of past diet*

Archaeogenetic study of ancient diets is an emerging area of research, with most studies performed within the last decade (Adler et al. 2013; Warinner et al. 2014; Weyrich et al. 2017). Many studies involve the direct sequencing of well-preserved archaeological flora or faunal remains. Comparing these sequences to modern lineages of the same specimens has allowed researchers to trace their domestication and adaptation (Pont et al. 2019; Irving-Pease et al. 2019). Some studies have used human paleogenomic data as a proxy to infer aspects of diet based on genetic markers. For example, the production of the enzyme lactase allows humans to digest lactose, the most abundant sugar in milk, and it is naturally produced in infancy. However, individuals with the lactase persistent (LP) phenotype have an extended lactase production that continues into adulthood (Gerbault et al. 2011). Until recently, there

was an assumption that individuals or populations that lacked this phenotype would not consume dairy (Ségurel and Bon 2017).

*1.1.4.a. Preservation of dietary DNA*

Compared to other biomolecules such as proteins, DNA is more easily degraded through cooking or exposure to taphonomic processes. However, the earliest genetic studies of ancient dental calculus demonstrated its extraordinary preservation potential (Adler et al. 2013; Warinner et al. 2014b, although potential dietary DNA was only a tiny fraction of the recovered sequences, and the vast majority (~99%) was identified as bacterial in origin (Warinner et al. 2014b; Weyrich et al. 2017). Investigations of dietary DNA within dental calculus to date have been limited, and many methodological issues concerning validation must be considered, outlined by Mann et al. (Mann et al. 2020).

*1.1.4.b. Limitations of paleogenetic studies of ancient diet*

While animal bones are abundant in the archaeological record, plant remains are much less common, and processes like carbonization that help preserve macroremains also hinder DNA preservation, making archaeogenetic studies of plants particularly challenging (Nistelberger et al. 2016). Additionally, the mere presence of macrobotanical or faunal remains within an archaeological site cannot be directly linked with human consumption. Although dietary DNA studies within dental calculus provide more direct evidence of consumption by individuals, humans often seek foods dense in starch or proteins. Lower yields of dietary DNA are expected for several reasons: (1) starch and protein-rich foods are not necessarily rich in DNA. Foods such as tubers have enlarged amyloplasts but not necessarily more cells or more DNA, (2) because tissues are largely made of proteins, foods have orders of magnitude more protein than DNA, (3) food preparation methods, such as cooking, have disproportionately negative effects on DNA stability compared to biomolecules such as protein, which contributes to degradation and fragmentation of DNA and (4) because calculus is a complex biofilm and all food biomolecules are likely to be present in small amounts compared to microbial biomolecules (Mann et al. 2020).

The reported numbers of dietary sequences present in proportions in some ancient calculus datasets fall within the reported levels of misidentification and carry-over contamination reported for Illumina sequencers (Warinner et al. 2014b; Sawafuji et al. 2020), and extra care must be taken to validate the authenticity of these sequences. Additionally, genomic analysis cannot provide tissue specificity, merely the identity of the consumed species. For instance, it would not be able to distinguish the consumption of cattle meat from cattle milk using DNA alone.

Many genetic dental calculus studies lack direct evidence of dietary consumption and instead use various proxies for diet. Researchers have used the microbial composition of dental calculus to infer shifts in dietary consumption (Adler et al. 2013; Ottoni et al. 2021). Others have inferred diet from the DNA of consumers through the presence of genes needed to produce various enzymes. However, there is always a risk of falsely equating functional capacity based on genes with actual human behavior. For instance, although the LP allele allows adults to digest lactose, it has become apparent it is not a prerequisite for dairy consumption, and a large body of evidence (including Manuscripts B and C in this thesis) shows that people without LP were practicing dairying as the management of domesticated bovids spread throughout southwest Asia, Europe, Africa, and Inner Asia (Burger et al. 2007; Dunne et al. 2012; Charlton 2019; Jeong et al. 2020; Bleasdale et al. 2021). While people with lactase non-persistence (LNP) can consume small amounts of lactose without significant

symptoms, milk products such as cheese, butter, and yogurt are processed by bacteria to convert lactose into lactic acid, effectively allowing LNP individuals to consume these dairy products in more significant quantities. Cultural adaptations and the use of technology is an important aspect of diet that must be considered.

Perhaps most importantly, the field is still very new and has not yet been fully established. Many methodological challenges make it especially difficult to validate dietary sequences within an overwhelmingly bacterial substrate. As with other forms of archaeological analysis, adequate databases are a concern. Some researchers have instead chosen to target sequences from specific species to confirm their presence and consumption. For example, Sawafuji et al. used rice (*Oryza sativa*) specific primers to confirm consumption (Sawafuji et al. 2020). However, these methods also suffer from the same issue discussed with other methods; the need for *a priori* knowledge precludes a more discovery-based analysis. In addition, PCR-based approaches cannot be used to authenticate ancient DNA findings. While there is great potential in paleogenomic dietary studies, there is also a clear need for more methodological work, and standard validation criteria should be adopted before the field can progress (Mann et al. 2020).

#### *1.1.5 Lipid residue studies of past diet*

Organic residue analysis of lipids from ceramic pottery has been instrumental in investigating ancient diet and subsistence. Researchers have demonstrated the high preservation potential of lipids preserved within the ceramic matrix of archaeological pottery, and the ability to extract and identify generalized categories of dietary lipids from ancient pottery is now well established (Evershed et al. 2008; Craig et al. 2011, 2012; Roffet-Salque et al. 2017; Dunne et al. 2020). It is particularly useful in the archaeological context due to the ubiquity of these artifacts worldwide.

##### *1.1.5.a. GC-MS and GC-C-IRMS*

Most archaeological lipid studies employ gas chromatography-mass spectrometry (GC-MS) or gas chromatography-combustion-isotope ratio mass spectrometry (GC-C-IRMS) techniques. The gas chromatography allows complex mixtures to be separated before the mass spectrometer identifies and quantifies the lipid biomarkers present in the sample based on the fragmentation patterns of individually identified peaks (single compounds). GC-MS is suitable for specific compounds with low molecular weights that are thermally stable and sufficiently volatile. Quantified fatty acids can be calculated into ratios used to identify general taxonomic origins based on reference ratios. As mentioned above, GC-C-IRMS is an analytical technique that employs both GC-MS and isotopic analyses. GC-MS analysis will first separate fatty acids before subsequent compound-specific isotopic analysis. For example, researchers have argued that the difference in $\delta^{13}C$ values of $C_{16}$ and $C_{18}$ ($\Delta 13C$) can be used to distinguish ruminant from non-ruminant fats (Copley et al. 2003).

##### *1.1.5.b. Preservation of dietary lipids*

Like other biomolecules, many complex factors can affect the preservation of lipids in the archaeological record (Evershed 2008), and preservation cannot be reliably predicted (Suryanarayan et al. 2021). The pottery matrix can allow for the binding of lipids within the ceramic pores, potentially trapping and preserving the molecules through time (Evershed 2008). However, like other artifacts, preservation can be affected by environmental factors such as soil pH and exposure to light and water. For example, researchers have demonstrated the dramatic effect of weathering on $C_{16:1}$ to $C_{:18:1}$ (Eerkens 2005) through time. Even before archaeological deposition, lipids can easily undergo chemical and structural modifications

from processes such as cooking.

*1.1.5.c. Limitations of organic residue analysis of pottery lipids*

While valuable, lipid analysis of pottery, which constitutes the vast majority of archaeological lipid studies, cannot provide direct evidence of dietary consumption on an individual or population level, merely a proxy for consumption. The presence of some food categories could even be associated with pottery production, such as the use of milk, fats, resins, or oils as post-firing sealants (Drieu et al. 2020). There is also a strict temporal limit to this analysis due to the relatively late adoption of pottery in human history, and pre-pottery populations will be inaccessible. Even when employing ceramic pottery technology, some populations may not process a particular dietary item within the vessels, and it would also remain invisible. Additionally, low-fat content dietary items are also likely to remain undiscovered using this technique. It is also important to consider the lack of taxonomic and tissue specificity inherent in the analysis. For example, the most common recovered dietary lipid compounds can only be classified into higher taxonomic categories such as "non-ruminant fat" or "plant wax" because the compounds could be derived from a large number of organisms within a dietary class.

Notably, the mixed-use of food types within the same vessel is possible and even likely, and problems of equifinality can complicate any interpretation of data to ascertain relative proportions of dietary sources (Fernandes et al. 2018). As described by Hendy et al., mixtures can lead to equivocal values that can be interpreted in multiple ways; mixing models demonstrate that when ruminant fat is mixed with $C_3$ plants, Δ13C values can resemble those of non-ruminant fats, and when $C_4$ plants are mixed with ruminants fed with $C_3$ crops, Δ13C values can resemble ruminant dairy fats (Hendy et al. 2018a). This problem has been encountered in subsequent lipid studies (Suryanarayan et al. 2021). Like other isotopic studies, isotopic baselines are vital for accurate interpretations, and previous knowledge about the local environment and available plant and animal species is vital for understanding factors that affect isotope values.

***1.1.6 Protein residue studies of past diet***

Studies in paleoproteomics have proliferated in the last decade due to technological advances in mass spectrometry methodologies. Unlike other biomolecules, proteins can be highly source-specific. Due to their relative abundance within living organisms, there is no need for artificial amplification where contamination can easily occur and be mistaken for the source material. Proteins have also been shown to survive better than the other biomolecules. The preservation limits are still unknown, but proteins have been successfully extracted from archaeological material several million years old (Rybczynski et al. 2013; Demarchi et al. 2016; Cappellini et al. 2019). Bone is the most studied source of ancient proteins, and collagen, in particular, has been the focus of much research. Collagen can be highly source-specific (Buckley et al. 2008), and past research has documented how resistant it is to various degradation processes, and many studies have successfully identified collagen within archaeological samples (Dobberstein et al. 2009). Bone and other materials, such as ceramics, are thought to protect proteins from biological and chemical degradation (Collins et al. 1995; Buckley and Wadsworth 2014). Due to their exceptional preservation, proteins have been utilized to show evolutionary relationships in skeletal remains without surviving DNA (Welker et al. 2015; Welker 2018; Chen et al. 2019). The two proteomic methods most frequently applied to archaeological material are peptide mass fingerprinting and tandem mass spectrometry.

*1.1.6.a. MALDI–time-of-flight (TOF) MS*

MALDI-TOF MS is used for peptide mass fingerprinting, a method used to separate a protein into smaller peptides. The mass/charge value is used to identify these peptides based on reference values, and if peptides are taxonomically diagnostic, a species can be assigned. Zooarchaeology by MS (ZooMS) is a method developed to identify the taxonomic origins of archaeological and historical material using MALDI-TOF MS (Buckley et al. 2009; Hendy 2021). It has been applied to many materials including, parchment (Kirby et al. 2013; Fiddyment et al. 2015; Teasdale et al. 2017), leather (Brandt and Mannering 2021), and eggshells but has most frequently been used to identify taxonomically distinct peptide masses from collagen within archaeological bone (Desmond et al. 2018; McGrath et al. 2019; Wagner et al. 2020; Korzow Richter et al. 2020; Martisius et al. 2020; Janzen et al. 2021). Biomarker identification through ZooMS can complement zooarchaeological studies when taxonomy cannot be assigned through morphology.

*1.1.6.b. Tandem Mass Spectrometry: LC-MS/MS*

Tandem mass spectrometry methods protein identifications are based on peptide fragmentation within the mass spectrometer, followed by the MS/MS analysis of the fragmented peptides, which provides the precise order of the amino acids within the peptide. This differs from peptide mass fingerprinting analysis, which gives the mass/charge (m/z) of peptides but does not further fragment the peptide, which allows for the determination of the specific amino acid sequences. The differences in sequence allow for the accurate identification of the peptides and taxonomic differences allow us to pinpoint the species of origin. This method also allows us to identify specific protein modifications and where they occur within a sequence (see Figure 1). Liquid chromatography-mass spectrometry (LC-MS/MS) technologies have become a popular approach for identifying ancient proteins due to their high sensitivity and accuracy. In theory, the ability to identify unknown compounds at

**Figure 1. Alignment of cow and sheep β-lactoglobulin (BLG).** Sequence differences are highlighted in yellow. MS/MS data of the C-terminal proteotypic peptide of BLG. This peptide has three sequence differences that can be used to distinguish (a) sheep from (b) cow

very low concentrations would make this instrumental approach ideally suited for the study of trace quantities of degraded protein residues one would expect to find in the archaeological record. While peptide mass fingerprinting via MALDI-TOF is appropriate for low complexity samples, or even single protein studies, LC-MS/MS is a more appropriate method for samples composed of taxonomically diverse and less predictable mixtures such as pottery or dental calculus. An important advantage of the method is that peptides identified are given probabilistic matches to known peptide sequences, which allows for statistical precision for characterizing protein residues. Shotgun, or full mass scan methods, have been applied to a wide variety of archaeological materials, food residues (Cappellini et al. 2010; Hong et al. 2012; Yang et al. 2014), paleofeces (Tsutaya et al. 2021), ceramics (Barker et al. 2015; Hendy et al. 2018a), textiles (Petroviciu et al. 2010; Ahmed et al. 2017; Li et al. 2021), and dental enamel (Parker et al. 2019a) and calculus (Warinner et al. 2014b, a; Hendy et al. 2018b; Jersie-Christensen et al. 2018; Wang et al. 2019; Geber et al. 2019; Charlton et al. 2019; Taylor et al. 2020; Wilkin et al. 2020; Scott et al. 2021; Bleasdale et al. 2021).

*1.1.6.c. Limitations of proteomic studies of ancient diet*

Depending on the substrate and burial context, ancient proteins can become highly modified over time. These changes can occur from cooking, diagenetic processes, or even from sample preparation. Post-translational modifications (PTMs) associated with proteomic degradation include deamination of glutamine or asparagine, methionine oxidation, and proline hydroxylation (Cappellini et al. 2014). These changes can lead to peptides with different molecular weights than the originals, leaving many ion spectra from ancient samples unassigned if not factored into data analyses.

Despite the reported success, the data-dependent proteomic methodologies employed in archaeological chemistry have limitations. Non-targeted (or data-dependent acquisition) mass spectrometry approaches can provide a general scan of protein residues within a sample but may not pinpoint archaeological dietary proteins of interest because endogenous and exogenous bacterial proteins can overwhelm their signal. Data acquisition is considered dependent because the most frequently identified peptides correspond to abundance, with higher abundance proteins being more likely to be identified (Michalski et al. 2011). This stochastic nature is a fundamental characteristic of shotgun proteomics. It means a bias inherent in the instrument towards repeatedly fragmenting the peptide sequences with the most intense MS signal, and less abundant peptides are less likely to be chosen for fragmentation and, therefore, will be much less likely to be identified. This can be especially problematic for archaeological samples, where soil, bacterial, or contaminant proteins are likely to overwhelm signals coming from lower abundance proteins of interest, such as dietary proteins. Using a targeted data-independent acquisition to identify only specified peptides can decrease the background noise that can overwhelm the signal of low abundance proteins. However, this requires *a priori* knowledge about which peptides are expected and would limit discovery-based findings

## 1.2 Dental calculus

***1.2.1 A reservoir of ancient biomolecule*s**

When dental plaque is not regularly removed, it builds up on the surface of teeth and spontaneously calcifies in layers throughout an individual's life. The plaque itself is a continuously forming, structured dental biofilm composed of diverse microbial communities and food and oral fluids particles. As food is chewed, dietary particles such as plant microremains and biomolecules such as DNA, proteins, lipids, and carbohydrates become

trapped in this sticky biofilm. The proteolytic metabolism of bacteria occurs within the gingival crevices of teeth, which leads to a localized increase in pH, which favors the mineralization of plaque. However, not all materials are broken down by proteolytic bacteria, and cellular components can become covered by a dental plaque within hours of eating. As plaque accumulates, it periodically calcifies and hardens into dental calculus, protecting biomolecules from further breakdown by salivary and bacterial enzymes, and embedding them within the calculus matrix. Plaque calcification can commence within a few days of colonization, but the timing varies widely (Lieverse 1999). As dental calculus forms, it accumulates into densely mineralized layers. These layers act as a reservoir, and various materials can become embedded within, collecting throughout a person's life. Biological material from human, bacterial, and dietary sources, as well as debris that may have passed through the mouth during a lifetime, either through accidental ingestion or inhalation, have been recovered from archaeological dental calculus samples (Radini et al. 2017). Although the precise mechanisms involved in the spontaneous calcification of dental calculus are still not fully understood, many complex factors play a role, including genetics, diet, mineral bioavailability, localized pH, and the surface features of any given tooth (Radini et al. 2017).

### *1.2.2 Advantages*

Before modern oral hygiene practices, the build-up of dental calculus was ubiquitous in the archaeological record. Dental calculus has been recovered within primates as far back in the archaeological record as the Miocene *Dryopithecus* (Fuss et al. 2018). Due to its excellent preservation potential, researchers can use dental calculus to study various temporal and geographic contexts. Within the last decade, the potential of this substrate has become apparent, and many studies have revealed the types of biological and microscopic data preserved within (Henry and Piperno 2008; Warinner et al. 2014b; Buckley et al. 2014; Hendy et al. 2018b; Radini et al. 2019). Due to its location, dental calculus can directly link an individual and digestion, which is particularly useful when investigating diet both on an individual and population level. However, not all foods that are consumed will be found in calculus. Although the recovery of protein residues found within human dental calculus can provide a crucial direct line of evidence about subsistence strategies utilized in the past, it cannot explain how often a dietary item was eaten or its relative importance to an individual's the overall diet.

### *1.2.3 Protein preservation within dental calculus*

As the field of paleoproteomics has progressed along with technological advances in mass spectrometry technology, more non-structural and non-fibrous dietary-specific proteins have been recovered from various substrates. Although the mechanisms of protein preservation of fibrous proteins such as collagen in bone can likely be attributed, at least in part, to structural robustness and the fact that they exist within a mineralized matrix, the exact mechanisms of ancient protein preservation, in general, remain elusive, even more so for proteins within dental calculus. While it is important to understand the expected protein preservation within ancient calculus, experimental studies are not always feasible. Experimental cooking and burial experiments have been conducted on pottery and animal bone to assess how various parameters affect preservation and recovery in the early stages of deposition (Barker et al. 2012, 2015, 2018). Additionally, many studies have assessed optimal extraction methods utilizing archaeological bone and pottery (Buckley et al. 2008; Barker et al. 2012). However, experimental studies on dental calculus are much less feasible for several reasons: (1) archaeological dental calculus is finite and exists in only small quantities, making replication studies impractical; (2) the proteomic approaches most appropriate for the complexity of samples such as calculus are destructive, and there is an ethical concern in exhausting these

materials for experimental studies; (3) recreating calculus as a realistic synthetic model has not yet been achieved; and (4) protein preservation within dental calculus has high inter-and intra- sample variability. Calculus formation varies widely at the individual and population level, and complete calculus sampling from individuals typically only results in the recovery of approximately tens of milligrams, although rarely, hundreds of milligrams can be obtained in some exceptional cases. Proteins are also not uniformly incorporated into calculus, although this is poorly studied, and more precisely defining this process would require a substantially large and homogenous sample of calculus with known dietary proteins to produce statistically valid experimental results. In reality, these criteria can rarely, if ever, be met in archaeological studies, hindering potential meaningful research in experimental studies. As a result, to better understand dietary protein preservation within archaeological dental calculus, we evaluated a number of characteristics that may play a role in the preservation of the proteins successfully recovered from both our dataset and from other paleoproteomic studies. Several shared characteristics have enabled us to form hypotheses about potential preservation mechanisms, which are discussed in section 8.2.

However, prior to the successful recovery of most of our dietary proteins, we made several changes to our working extraction protocol. While the focus of this thesis is to interpret identified dietary proteins from archaeological dental calculus, in order to reconstruct the dietary features of specific ancient societies, a vital part of this research also included the development of more optimized protein extraction protocols. At the onset of this thesis project, although early milk protein recovery was low but relatively consistent, the extraction protocol in use appeared to be inefficient, with high overall performance variability. However, due to the challenges previously described above, it was difficult to determine whether the low and variable recovery of dietary proteins from ancient dental calculus was due to inherent preservation variability or due to the variable performance of the extraction methods. In order to try and improve overall performance and consistency, the protocol was evaluated for areas of potential optimization, and several changes were made, summarized in Table 2. Multiple steps in the original protocol were determined to likely lead to decreased performance and digestion efficiency. The changes to the protocol and the reasoning behind each of the changes are discussed in section 1.3.1.b.

After these protocol changes, culminating in the protein identifications in Manuscript C, we had extraordinarily high levels of milk peptide recovery compared to previous archaeological studies- although whether or not this can be attributed to improved methodological effectiveness or the high level of consumption and preservation of the region cannot be assessed. However, the high level of non-dairy peptides recovered from Bronze and Iron Age Levant is far beyond the levels of previous non-dairy dietary protein recovery from dental calculus, which typically consisted of a hand full of peptides. The subsequent sample sets we have applied the updated methods to are consistently yielding higher numbers of total protein identifications, including previously unidentified types of proteins from dental calculus, and it is clear that method optimization is improving overall protein recovery and identification. This will be discussed further in the Discussion.

## 1.3 Methods used in this thesis

A filter aided sample preparation (FASP) approach was used to extract proteins for all four manuscripts described in this thesis. The initial FASP protocol used at the onset of this project (Version 1) was used for samples included in Manuscript A of this thesis as well as other protein studies (Jeong et al. 2018; Bleasdale et al. 2020; Wilkin et al. 2020). An optimized version of this protocol (Version 2) was applied in the subsequent studies

presented in this thesis, as described below for Manuscripts B, C, and D.

### *1.3.1 Extraction protocols*

*1.3.1. a. Initial extraction protocol: Version 1*

Dental calculus samples were placed in 1.5 mL Surelock Eppendorf tubes and decalcified in 1 mL 0.5M ethylenediaminetetraacetic acid (EDTA) on a rotator at room temperature for 3-7 days. Samples were then centrifuged to pellet the insoluble materials. The supernatant was removed, and 800 mL were placed in -80°C storage. The remaining 200mL of supernatant was added to 50 uL of 8M urea (UA), filtered in a 30 kD microcon filter, and centrifuged at 35°C for 20 minutes at 14,000 g.

The protein pellet was resuspended with 30 uL of sodium dodecyl sulfate (SDS) extraction buffer. The extraction solution stock was prepared with 90 uL of 20% SDS, 45 uL of 2.5M dithiothreitol (DTT), 45 uL of 1M Trizma pH 8.2, and 270 uL of Milli-Q water. Pellet samples were then heated for 5 minutes at 95°C, followed by centrifugation at 14,000 g for 5 minutes. The supernatant was transferred to the same microcon filter used for the EDTA fraction, along with 200 uL of 8M urea, and centrifuged at 35°C for 20 minutes at 14,000 g. Another 200 uL of 8M urea was added, and the samples were centrifuged for an additional 20 minutes. Flow-through was discarded. Alkylation followed, using 100 uL of 0.05M iodoacetamide (IAA)/UA solution. The samples were mixed for 1 minute in the dark at 600 rpm and incubated for an additional 5 minutes. Samples were spun for 12-15 minutes at 35°C (14,000 g) and washed twice with 100 uL of 8.0M UA. Flow-through was discarded. The sample was then washed twice with 100 uL of 0.5M sodium chloride (NaCl). Filter units were transferred to a new collection tube, and 120 uL of a 0.01ug/uL solution of trypsin was added to the filter and allowed to digest overnight at room temperature. Samples were centrifuged he next day at 14,000 g for 15-20 minutes. Peptides were collected in the new collection tube and acidified in a 5% trifluoroacetic acid (TFA) solution to stop the digestion process.

*1.3.1.b. Updated protocol: Version 2*

Dental calculus samples were placed in 1.5 mL Surelock Eppendorf tubes and decalcified in 500 uL 0.5M EDTA on a rotator at room temperature for 3-7 days. This decrease of EDTA from 1000 uL to 500 uL ensured a more concentrated supernatant. 200 uL of the supernatant was filtered through a 30kD microcon filter in two steps, washed with 200 uL of 8.0M urea, and centrifuged at 18°C for 20 minutes (14,000 g). The remaining 300 uL were removed for backup and placed in a -80°C freezer. The temperature during all subsequent spins was also set to 18°C, reducing the 35°C of the previous protocol. Over time, or in the presence of high temperatures, urea can increase the likelihood of carbamylation of the N-termini of proteins.

**Table 2. Summary of differences in extraction methods used in this thesis**

| | **Version 1** | **Version 2** |
|---|---|---|
| EDTA for decalcification | 1000 mL | 500 uL |
| Temp. for centrifugation | 35°C | 18°C |
| DTT | 2.5M | 1.0M |
| Alkylation step | 5 minutes | 20 minutes |
| UA washes after alkylation | 2x 100 uL of 8.0M UA | 2x 100 uL of 8.0M UA |
| Wash step prior to digestion | 2x 100 uL of 0.5M NaCl | 3x 100 uL of 0.05M TEAB |
| Digestion conditions | Room temperature | 37°C in thermomixer at 300 rpm |
| Post digestion washes | None | 2x 40 uL TEAB, 1x 50 uL NaCl |

Carbamylation of Lysine and Arginine residues can also occur, preventing proteolytic cleavage of these residues by the enzyme trypsin, which hinders identification by LC-MS. In reality, urea-induced carbamylation is a common problem; many currently published proteomic datasets show this modification (Kollipara and Zahedi 2013). Wisniewski et al. address this problem by suggesting that urea solutions are made fresh before use and that all FASP steps are carried out at room temperature (18-22°C) (Wiśniewski et al. 2009).

As in the previous protocol, 30 uL of SDS solution was added to the protein pellet and then the samples were heated for 5 minutes at 95°C. The digestion buffer was the same as the previous protocol, but the DTT was reduced from 2.5M to 1.0M. DTT is used to reduce disulfide bonds between cysteine residues, which play a substantial role in maintaining protein stability, and a chemical reducing agent is needed to ensure the proteins can be denatured and solubilized. 1M DTT is standard and adequate for our protein extractions, and a proper ratio of denaturing to alkylating reagents is needed to prevent a disruption to either process. Following heat denaturation, samples were centrifuged at 14,000 g for 10 minutes, and the resulting supernatant containing proteins was transferred to the same microcon filter used for the EDTA fraction. 200 uL of 8M urea was used to wash the sample and centrifuged at 18°C for 20 minutes at 14,000 g. Another 200 uL of 8M urea was added, and the samples were centrifuged for an additional 20 minutes. The flow-through was discarded.

Alkylation of the samples with IAA followed to prevent disulfide bonds from reforming after reduction by DTT. 100 uL of 0.05M IAA/UA solution was allowed to incubate in the dark for 20 minutes, increasing from the previous protocol. This increased time allows the alkylation process more time to proceed. The decrease in DTT, which consumes IAA, decreases the chance of under-alkylation and cysteine bond reformation, which would otherwise hinder trypsin digestion and subsequent protein identification. After incubation, samples were spun at 18˚C for 12-15 minutes at 14,000 g and then washed with 200 uL of 8.0M urea. This step was repeated two more times. This increase in the number of washes from the previous method ensured adequate removal of SDS, which can hinder trypsin digestion and LC-MS analysis and cause damage to instrumentation (León et al. 2013). Unlike the previous protocol, no NaCl washes were performed. Both NaCl and urea can have a concentration-dependent negative effect on trypsin digestion efficiency. Instead, buffer exchange via three washes of triethylammonium bicarbonate (TEAB) was performed to ensure the proper removal of urea. 120 uL of a 0.01 ug/uL solution of trypsin was added to the filter and incubated overnight at 37°C in a thermomixer at 300 rpm. The previous protocol specified a room temperature digestion overnight. Modified trypsin, which reduces autolysis, was utilized in all experiments. Autolysis of trypsin is a common problem in proteomics, as trypsin itself has 14 lysyl and two arginyl residues within its sequence that are potential attack sites for self-hydrolysis. Autolysis can be reduced or even eliminated with the use of trypsin that has been modified by reductive methylation. Modified trypsin is common in many proteomics laboratories for this reason. However, Havis et al. demonstrated that reductive methylation shifts trypsin's optimum catalytic activity to between 50 and 60°C (Havlis et al. 2003). This increased temperature also led to a digestion efficiency that was twelve times faster than the typical 37°C digestion used by many labs. Michaelis-Menten kinetic parameters have been established to the modified trypsin currently used in our protocol (Promega). Higher temperatures increase the collision energy of the molecules, and room temperature digestion is much less effective for modified trypsin. We set our digestion temperatures to 37°C due to the timing of the protocol, which requires overnight digestion for a convenient stopping place for a two-day protocol.

The next day, the filter units were transferred to new collection tubes. Unlike the previous protocol, after digestion, three additional centrifugation steps using TEAB and NaCl were added to this protocol to increase the elution and recovery of peptides from the microcon filter unit. First, 40 uL of TEAB solution was added before samples were centrifuged at 14,000 g for 15-20 minutes. This step was then repeated. Next, 50 uL of 0.5M NaCl solution was added, and the samples were spun once more for 10 minutes. To stop the digestion process, the filtrate containing the peptides was acidified in a 5% (TFA) solution to a pH below 3.

*1.3.1.c. Desalting*

C-18 Empore (3M) solid-phase extraction (SPE) StageTips were prepared in-house or purchased (ThermoScientific). The StageTips were cleaned with 150 uL methanol, conditioned with 150 uL 60% acetonitrile (ACN)/0.1% TFA solution, and then equilibrated twice with 150 uL of 3% ACN/0.1% TFA solution. All spins were performed for 1 minute at 2000 g. Extracted peptides were then loaded into the StageTips and centrifuged. Samples were then washed twice with 3% ACN/0.1% TFA and samples were eluted into a new tube with two rounds of 60% ACN/0.1% TFA solution. The solution was then dried to completeness in a centrifugal evaporator and stored in a -80°Cfreezer until they were shipped on dry ice to the Functional Genomics Center in Zurich, where the dried samples were then resuspended in 15 uL of 3% ACN/0.1% TFA.

### *1.3.2 LC-MS/MS*

Samples were analyzed by LC-MS/MS using either a Q-Exactive HF (Manuscript A) or a Q-Exactive mass spectrometer (Manuscripts B, C, and D) (Thermo Scientific, Bremen, Germany) coupled to an ACQUITY UPLC M-Class system (Waters AG, Baden-Dättwil, Switzerland) at the Functional Genomics Center Zurich of the University/ETH Zurich through full mass scans.

For the Q-Exactive HF mass spectrometer, spectra were acquired from 350-1500 m/z with an automatic gain control target of 3e6, a resolution of 120,000 (at 200 m/z), and a maximum injection time of 50 ms. The quadrupole isolated precursor ions with a 1.2 m/z window, a 1e5 automated gain control value, and a maximum fill time of 50 ms. Twelve of the most intense precursor ions for each MS1 scan were fragmented with a normalized collision energy of 28 and scanned with a resolution of 30,000 (at 200 m/z), and 130 m/z specified as a fixed first mass . An intensity threshold of 9e4 was applied for MS2 selection, and singly charged ions were excluded. Filter criteria for MS2 selection included an intensity threshold of 9e4, and unassigned, singly charged ions or ions with a charge state greater than eight were excluded. Selected precursor ions were put onto a dynamic exclusion list for 30 seconds.

For liquid chromatography, the solvent composition at the two channels was 0.1% formic acid (FA) in $H_2O$ for channel A, and 0.1% formic acid in acetonitrile for channel B. 4 µL for each peptide sample were loaded to a trap column (Symmetry C18, 100 Å, 5 µm, 180 µm x 20 mm, Waters AG, Baden-Dättwil, Switzerland) with a flow rate of 15 µl/min of 99 % solvent A for 30 seconds at room temperature. Peptides eluting from the trap column were refocused and separated on a C18 column (HSS T3 C18, 100 Å, 1.8 µm, 75 µm x 250 mm, Waters AG, Baden-Dättwil, Switzerland). The column temperature was 50 °C. After 1.5 min gradient stabilization at 99 % solvent A, peptides were separated over 120 minutes with the following gradient: 1% - 5 % solvent B in 30 seconds, 5% - 40% solvent B in 120 min. The column was cleaned with 98 % solvent B for 5 min after the separation and reequilibrated at loading condition for 8 min before initializing the next run. Potential contamination and

sample carryover were monitored through extraction blanks and injection blanks between each sample. The parameters for the Q-Exactive mass spectrometer were the same, with a few exceptions, highlighted in Table 3

### *1.3.3 Data Analysis*

All samples in this thesis were analyzed using Mascot (Matrix Science, London, UK). Raw tandem mass spectra were converted to Mascot generic files by MSConvert, using the 100 most intense peaks in each spectrum. Mascot was set up to either search the databases SwissProt or SwissProt and TrEMBL together. All searches were performed assuming the digestion enzyme trypsin and selecting the automatic decoy option. Mascot search parameters included a fragment ion mass tolerance of either 0.01 Da or 0.05 Da, depending on the LC-MS instrumentation used. A parent ion tolerance of 10 ppm was specified for each sample. Carbamidomethyl (C) was always specified as a fixed modification due to the use of IAA during protein extraction. Oxidation of methionine, deamidation of asparagine, and deamidation of glutamine were also specified as variable modifications. Scaffold (Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications.

Datasets were filtered, so that protein false discovery rate (FDR) was less than at least 5%, and peptide FDR was less than at least 1%, which required a 2-3 peptide minimum for protein identifications for each sample. Protein identifications also required a minimum of two unique identified peptides. Proteins that could not be differentiated based on MS/MS analysis alone were grouped into clusters to satisfy the principles of parsimony. Data files for each project were uploaded to the PRIDE repository under the identifiers PXD015002, PDX027706, PDX027728, and PXD021498. Dietary proteins were manually validated by searching all peptides against the NCBi nr database using BLASTp to assess taxonomic specificity

**Table 3. Summary changes in of instrument settings between the Q-Exactive HF and the Q-ExactiveF**

| | Q-Exactive HF | Q-Exactive |
|---|---|---|
| Orbitrap resolution | 70,000 (200 m/z) | 70,000 (200 m/z) |
| Maximum injection time | 50 ms | 110 ms |
| Quadrupole scan window | 1.2 m/z | 2.0 m/z |
| Automated gain control value | 1e5 | 5e4 |
| Maximum fill time | 50 ms | 110 ms |
| Normalized collision energy | 28 | 25 |
| Precursor ion scan resolution | 30,000 | 35,000 |
| First fixed mass | 130 m/z | 200 m/z |
| $MS_2$ intensity threshold | 9e4 | 1.9e3 |
| 99% Solvent A column load time | 30 seconds | 60 seconds |
| Run time | 120 minutes | 75 minutes |
| Gradient: Solvent B | 1-5% in 30 seconds | 8-22% in 49 seconds |
| | 5-40% in 120 minutes | 22-32% in 11 minutes |
| Column cleaning | 99% Solvent B | 95% Solvent B |
| Re-equilibration time | 8 minutes | 10 minutes |

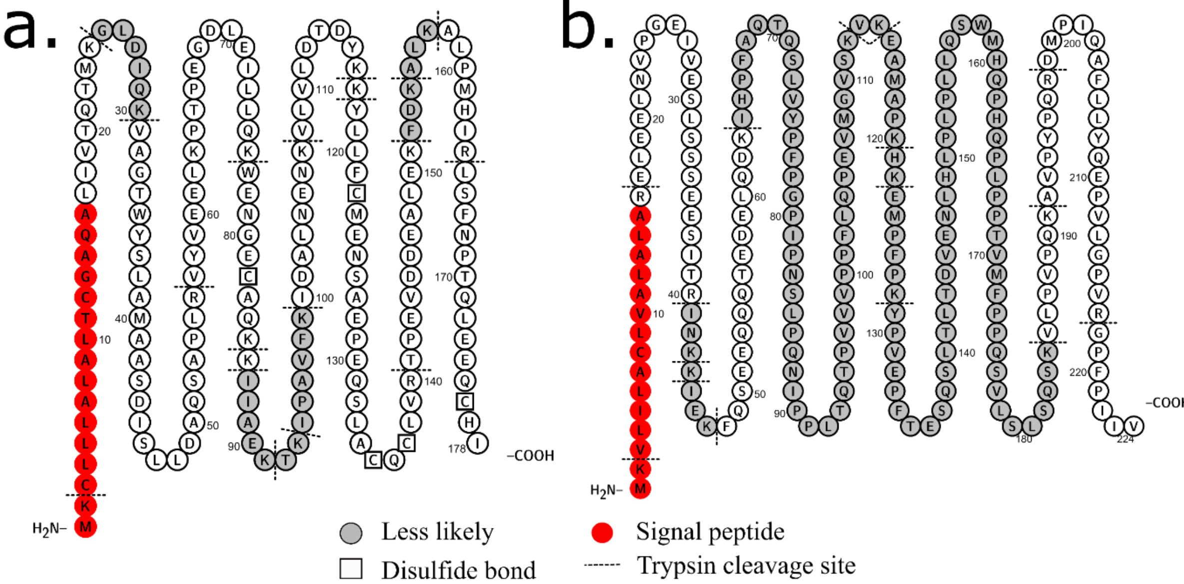



**Figure 2. Proteotypic recovery of peptides** (a) bovine BLG and (b) bovine β-casein. Signal peptides, shaded in red, are not typically recovered. Peptides shaded in gray are less likely to be recovered by LC-MS/MS due to their length or hydrophobicity (Bislev et al. 2012). This figure was made using the protein sequences downloaded from Protter (Omasits et al. 2013)

Requiring a minimum threshold for coverage may be feasible for many modern protein studies or paleoproteomic studies of very large proteins such as well-preserved collagen. However, this is less feasible for dental calculus, as many of our dietary proteins of interest have limited proteotypic peptides and imposing strict minimums could make them impossible to detect. We do not observe all peptides with equal probability, and many factors contribute to the detectability of any given peptide. Some peptides are preferentially identified, while others only show up rarely or not at all. A protease (typically trypsin) will cleave a protein at specific sites during digestion resulting in a subset of tryptic peptides. Of these peptides, only some will be successfully ionized and be detected. These "flyable" peptides are most likely to be detectable in any given run. A subset of these peptides known as "proteotypic peptides" are those peptides that are the most likely to be confidently and consistently observed by current MS-based proteomics methods (Craig et al. 2005), based on a number of physicochemical factors. Peptides are more likely to be identified when they are of a specific length (typically between 8 and 25 amino acids) and not too hydrophobic or hydrophilic (Jarnuczak et al. 2016). These proteotypic peptides are simply *easier* to identify. Because these peptides are readily identifiable, the most frequently observed peptides do not necessarily correspond to the most abundant proteins within a sample. For example, although the casein component of milk accounts for approximately 80% of the protein content in cattle milk, compared to the approximately 20% of whey proteins, BLG is the most commonly recovered protein in the archaeological record. In contrast, the recovery of caseins is much lower. The higher detection rate of BLG is likely due, at least in part, to BLG's large number of proteotypic peptides compared to caseins, illustrated in (Figure 2). Additionally, unoptimized methods can result in incomplete digestion or missed cleavages, leading to a decrease in the identification rates of some peptides. During shotgun proteomics experiments, intense MS signals from abundant peptides can drown out the less abundant peptides, and peptide ion co-elution can affect the ionization and detectability of many peptides (Jarnuczak et al. 2016).

### *1.3.4 Dairy Database*

The choice of database influences the ability to detect proteins and peptides of interest. Proteins not present in the database of choice cannot be found even if they are contained within a sample. SwissProt, the annotated portion of Uniprot, is a database that is often used in proteomic research. This database is both curated and non-redundant, which decreases search times dramatically. While SwissProt is large enough to allow for the calculation of statistical significance of results, it lacks many proteins from traditional dairying livestock species that are of interest in our dairy focused proteomic studies, including sheep, horse, zebu, donkey, camel, yak, and reindeer (see Figure 3). While TrEMBL, the unannotated and redundant portion of Uniprot contains more milk protein sequences, it also dramatically increases our search space and seach times, as well technological needs such as higher CPU or memory. As of June 2021, SwissProt contains 565,254 sequence entries, while TrEMBL contains 219,174,961 sequences, 99.3% of which are either inferred from homology and predicted (UniProt Consortium 2021).

In order to assess if we could increase milk protein identification to correct any potential biases in our data by searching only SwissProt, while avoiding substantially increased search times, we compiled a custom database that can be searched alongside SwissProt during data analysis. Ten of the most abundant and taxonomically informative proteins from 12 species were chosen, including alpha $\alpha S_1$ casein ($AS_1$-CN), $\alpha$-$S_2$- casein ($AS_2$-CN), β-casein (B-CN), κ-casein (κ-CN), β-lactoglobulin (I) (BLG), β-lactoglobulin II (BLG II), α lactalbumin (A-LAC), serum albumin (SA), lactotransferrin (LTF), and lysozyme C (LYZC). Only two livestock species (cattle and sheep) were fully represented in SwissProt for the 10 proteins listed above, shown in Figure 3. Sequence data for the remaining proteins were collected from TrEMBL or translated from NCBI GenBank and were compiled into a database. However, we got comparable results from using SwissProt and SwissProt plus the Dairy database , and to date, none of the additional dairy proteins have been identified in samples discussed in this thesis, and, therefore, this additional database was not included in the published articles discussed within. This may be because only a small number of milk proteins persist in dental calculus, and these are well represented in SwissProt.

| | $AS_1$-CN | AS-CN | B-CN | K-CN | BLG-(I) | BLG-(II) | A-LAC | SA | LTF | LYZ C |
|---|---|---|---|---|---|---|---|---|---|---|
| *Bubalus bubalis* | | | | | | | | | | |
| *Bos grunniens* | | | | | | | | | | |
| *Bos taurus indicus* | | | | | | | | | | |
| *Bos taurus* | | | | | | | | | | |
| *Capra aegagrus hircus* | | | | | | | | | | |
| *Ovis aries* | | | | | | | | | | |
| *Rangifer tarandus* | | | | | | | | | | |
| *Camelus dromedarius* | | | | | | | | | | |
| *Camelus bactrianus* | | | | | | | | | | |
| *Equus ferus caballus* | | | | | | | | | | |
| *Equus africanus asinus* | | | | | | | | | | |
| *Homo sapiens* | | | | | | | | | | |

| **SwissProt** | **TrEMBL only** | **NCBi only** | **Not available** | **Not present** |
|---|---|---|---|---|

**Figure 3. Database coverage of milk proteins.** Many milk proteins from traditional livestock species were unavailable in SwissProt. Yellow indicates proteins that were annotated and available through SwissProt searches. Light gray indicates proteins that were available in the unannotated TrEMBL database. Dark gray represents proteins not available in either portion of Uniprot but were available in NCBi's GenBank and were translated into protein sequences. Red denotes proteins that were not publicly available in any of the previously mentioned sources, while black denotes proteins not present in that particular species.

# 2. Aims of the Thesis

This thesis aims to expand upon previous research into the recovery and identification of dietary proteins from the archaeological record. The overarching goal is to identify and verify dietary protein residues, taxonomically distinctive, or even tissue level to characterize more clearly what people ate in the past. The ability to extract and characterize dietary proteins from dental calculus could enable researchers to study subsistence in ways unavailable with other methods. However, to date, the recovery of taxonomically distinct unequivocal dietary proteins beyond collagen and dairy is rare.

Considering this, the manuscripts presented here serve three important goals: (i) as a proof-of-concept for our ability to recover more diverse dietary proteins from the archaeological record, (ii) as a way to determine how proteomic methods complement other archaeological methods, and in which contexts proteomic methods would be the most useful, and (iii) providing an opportunity to look for patterns in the types of dietary proteins recovered from dental calculus to better assess what kind of biases are introduced in our dietary studies.

We applied proteomic methods in four different studies discussed in this thesis. The samples originated from a wide range of archaeological contexts and anticipated dietary customs, including from:

1. Individuals who perished at a $19^{th}$ century rural Irish workhouse for the poor during the Irish Famine. These individuals were provided relief rations in the time leading up to their deaths, and historical accounts give us some clues about the content of these diets.

2. Individuals from the Bronze Age Xiaohe culture in the Tarim Basin in northwest China that practiced both agriculture and pastoralism near a lake environment within a harsh and dry desert environment.

3. Individuals that span the Eneolithic through the Roman Era in the Northern Caucasus and surrounding areas. These individuals include some of the earliest known inhabitants of the Northern Caucasus, whose subsistence economy was not fully known. Later Bronze Age steppe populations in the region were believed to be the first fully mobile pastoralists, a subsistence that is almost entirely focused on domesticated dairy livestock.

4. Individuals from Bronze and Iron Age southern Levant who lived in large urban settlements and did not practice extreme forms of subsistence, but instead likely relied on staple agricultural plants, and domesticated animals. Additionally, there was a demand for exotic goods from other regions in these urban populations and they were known to engage in long-distance trade, although the full extent of the trade in dietary items is not fully understood.

A number of additional research questions were addressed in each study, listed below:

**Manuscript A:**

- Can molecular techniques augment historical data, or will our results align with expected dietary items?
- Can microparticle and proteomic analyses elucidate aspects of food availability and

consumption patterns amongst a historical population?

**Manuscript B:**

- Were the earliest inhabitants of Xiaohe already consuming milk, or was milking technology established after the population settled in the region through the spread of dairy pastoralism from nearby populations?
- Can we establish which dairy species they relied on?
- How prevalent was dairy consumption?
- What insights can dietary proteomic studies provide in the context of studies involving population genetics?

**Manuscript C:**

- Were individuals in the Northern Caucasus practicing dairy pastoralism?
- When did dairy technology first arrive in the region?
- How prevalent was the use of dairy technology?
- Which species were utilized for dairy consumption?
- Do we see differences through time, between cultural groups, or in ecotones?
- What implications does dairy pastoralism in the Northern Caucasus have for the Bronze Age steppe herder migrations?

**Manuscript D:**

- Can molecular methods reveal elements of trade and cuisine that otherwise leave few archaeological traces?
- What is the overall protein recovery from Bronze Age samples from the Southern Levant?
- Did long-distance trade facilitate these diets?
- How complimentary are proteomic methods to other methods used to study dietary

# 3. Overview of manuscripts and author contributions

## 3.1. Manuscript A

**Relief food subsistence revealed by microparticle and proteomic analyses of dental calculus from victims of the Great Irish Famine**

Johnny Geber*, Monica Tromp*, Ashley Scott*, Abigail Bouwman, Paolo Nanini, Jonas Grossman, Jessica Hendy, and Christina Warinner

**Published in *PNAS* (2019)**

*Johnny Geber, Monica Tromp, and Ashley Scott contributed equally to this work.

**Synopsis**:
In Manuscript A, we focused on molecular analyses of human dental calculus from individuals who died in the Kilkenny Union workhouse during the Irish famine. This project aimed to test for the presence of dietary remains to elucidate the variability of food provisions and how well these remains corroborated historical accounts. This project included microremain and proteomic components. Our results provide direct evidence of the consumption of various foods, including oats, potatoes, corn, wheat, egg, and milk, by individuals who took refuge in the workhouse.

We found preserved microremains in 41 of 42 analyzed samples and dietary proteins in 10 of 14 samples analyzed. Microparticles identified included those specific to corn, barley, wheat, oat, and potato. The milk protein β-lactoglobulin was recovered from seven individuals. Additionally, ovalbumin proteins were recovered from three individuals. Both the microparticle and proteomic results are consistent with the historical account of 19$^{th}$-century Irish laborers' diet and relief food consumed during the Famine. Calculus build-up in the analyzed sub-sample reflects tartar accumulations both before and during life in the workhouse. Potato, milk, and corn are all well-documented aspects of food rations from Irish workhouses. Corn, also known as Indian meal, was sent as relief from the United States and these results likely reflected consumption from the food provided in the workhouses prior to death. Proteins belonging to chicken eggs were also recovered, but these were not a common aspect of food relief during the famine and may reflect the diet of the individuals before the famine.

**Author contributions:**
Johnny Geber and Monica Tromp conceived and designed this research study, which consisted of the historical, microparticle, and proteomic components. Johnny Geber contributed to the historical research of the study. Monica Tromp provided the microparticle analysis, while Ashley Scott provided the analysis of the proteomic portion of the study. Monica Tromp analyzed a total of 42 samples, while Ashley Scott processed a total of 14 samples. Extracted samples were processed by Paolo Nanni and Jonas Grossman using the Q Exactive HF mass spectrometer at the Functional Genomics center of Zurich. Ashley Scott performed the ancient protein data analysis and verification with feedback from Paolo Nanni, Jonas Grossman, Jessica Hendy, and Christina Warinner. Abigail Bouwman and Christina Warinner contributed new reagents/analytical tools. Johnny Geber, Monica Tromp, and Ashley Scott composed the manuscript contributions from Jessica Hendy and Christina

Warinner.

| | |
|---|---|
| Sample procurement, organization, and processing | 20% |
| Ancient protein laboratory work | 0% |
| Ancient protein bioinformatic processing | 95% |
| Ancient protein verification and quality control | 95% |
| Manuscript Construction and Organization | 30% |

Ashley Scott contributed 30% to the overall project, including laboratory work, data analysis, and manuscript construction.

Note: Supplementary Data files can be accessed via the following DOI: https://doi.org/10.1073/pnas.1908839116

Jena, 11.10.2021

______________________________

Ashley Scott

## 3.2 Manuscript B

**The genomic origins of the Bronze Age Tarim Basin mummies**

Fan Zhang*, Chao Ning*, Ashley Scott*, Qiaomei Fu, Rasmus Bjørn, Wenying Li, Dong Wei, Wenjun Wang, Linyuan Fan, Idilisi Abuduresule, Xingjun Hu, Qiurong Ruan, Alipujiang Niyazi, Guanghui Dong, Peng Cao, Feng Liu, Qingyan Dai, Xiaotian Feng, Ruowei Yang, Zihua Tang, Pengcheng Ma, Chunxiang Li, Shizhu Gao, Yang Xu, Sihao Wu, Shaoqing Wen, Hong Zhu, Hui Zhou, Martine Robeets, Vikas Kumar, Johannes Krause, Christina Warinner, Choongwon Jeong, Yinqiu Cui

**Accepted to *Nature* (July 2021), in press**

*Zhang, Ning, and Scott contributed equally to this study

**Synopsis:**
In Manuscript B, we analyzed the remains of Xinjiang's earliest inhabitants discovered thus far, dating from 3000-1700 BCE. We presented the genomic data from 5 individuals from the Dzungarian Basin in north Xinjiang and 13 individuals from the Tarim Basin in south Xinjiang. Our results show that the inhabitants of the Dzungarian Basin, dating to ca. 3000-2800 BCE, exhibit an Afanasievo ancestry with an additional local contribution, while the Tarim Basin individuals, dating to ca. 2100-1700BCE, only harbor a local ancestry. Additionally, we present protein analysis data from the dental calculus of n=7 individuals from the earliest layers at the site of Xiaohe in the Tarim Basin. All seven individuals exhibited strong evidence of milk proteins in their dental calculus, indicating a reliance on dairy pastoralism at the site since its founding.

Our results do not support previous hypotheses for the origin of the Tarim mummies, who were argued to be descended from either the Afanasievo, BMAC, or IAMC cultures. Instead, we find that the earliest Tarim Basin cultures appear to have arisen from a genetically isolated local population that adopted neighboring pastoralist and agriculturalist practices, which allowed them to settle and thrive along the shifting riverine oases of the Taklamakan Desert.

**Author contributions:**
This study was conceived and supervised by Yinqiu Cui, Choongwon Jeong, Christina Warinner, Chao Ning, and Johannes Krause. It included both a genetic and protein analysis component as well as linguistic discussions. Ashley Scott's contribution to the project included the protein analysis portion. Ashley Scott performed the protein extractions in the dedicated ancient protein clean lab at the Max Planck Institute for the Science of Human History in Jena, with assistance from Linyuan Fang whom Ashley Scott was training. Each sample was cataloged and weighed prior to analysis. Only individuals with calculus deposits >5 mg were analyzed, and 5-10 mg of dental calculus was processed for each sample. In total, Ashley Scott processed dental calculus samples from seven individuals, some in duplicate for a total of 18 samples. Ashley Scott analyzed and authenticated the protein data with feedback from Christina Warinner. For the genomic portion of the study, Fan Zhang, Peng Cao, Ruowei Yang, Feng Liu, and Qingyan Dai performed laboratory work while Chao Ning, Fan Zhang, Christina Warinner, Choongwon Jeong, Qiaomei Fu, Pengcheng Ma, Xiaotian Feng, Wenjun Wang, and Vikas Kumar analyzed the DNA data. Qiaomei Fu, Dong Wei, Wenying Li, X Xingjun Hu, Qiurong Ruan, Idilisi Abuduresule, Chunxiang Li, Shizhu

Gao, Yang Xu, Shaoqing Wen, Hong Zhou, and Alipujiang Niyazi provided archaeological materials and associated information. Rasmus Bjørn and Martine Robeets provided the linguistic background, and Guanghui Dong and Zihua Tang helped with AMS datings. Chao Ning, Fan Zhang, and Ashley Scott wrote the manuscript with significant contributions from Christina Warinner, Choonwon Jeong, and Yinqiu Cui, and with input from all coauthors. Ashley Scott's contribution to the manuscript focused primarily on the protein analysis portion of the study.

| | |
|---|---|
| Sample procurement, organization, and processing | 10% |
| Ancient protein laboratory work | 100% |
| Ancient protein bioinformatic processing | 100% |
| Ancient protein verification and quality control | 100% |
| Manuscript Construction and Organization | 10% |

Ashley Scott contributed 30% to the overall project, including laboratory work, data analysis, and manuscript construction.

Note: Supplementary Data files can be accessed via the following DOI: https://doi.org/10.1038/s41586-021-04052-7

Jena, 11.10.2021

________________________________

Ashley Scott

## 3.3. Manuscript C

" **Emergence and intensification of dairying in the Caucasus and Eurasian steppes** "

Ashley Scott, Sabine Reinhold, Taylor Hermes, Alexey A. Kalmykov, Andrey Belinskiy, Alexandra Buzhilova, Natalia Berezina, Anatoliy R. Kantorovich, Vladimir E. Maslov, Farhad Guliyev, Bertille Lyonnet, Parviz Gasimov, Tufan Axhundov, Bakhtiyar Jalilov, Jeyhun Eminli, Emil Iskandarov, Emily Hammer, Selin Nugent, Richard Hagan, Kerttu Majander, Päivi Onkamo, Kerkko Nordqvist, Natalia Shishlina, Elena Kaverzneva, Arkadiy I. Korolev, Aleksandr A. Khokhlov, Roman V. Smolyaninov, Rüdiger Krause, Eliza Stolarzyk, Maria Karapetian, Svetlana V. Sharapova, Johannes Krause, Svend Hansen, Wolfgang Haak, Christina Warinner

**Published in *Nature Evolution and Ecology*, June 2022**

**Synopsis**:

In Manuscript C, we applied high-resolution proteomic methods to investigate the dietary foundations of individuals through the proteomes of dental calculus from humans buried in the Northern Caucasus and the surrounding Southern Caucasus Oka-Don-Volga, and Ural regions. Samples range in date from the Neolithic through the Roman era. Current archaeological and archaeogenetic evidence points to the Pontic-Caspian steppe zone between the Caucasus and the Black Sea as the crucible from which the earliest steppe pastoralist societies arose and spread, ultimately influencing populations from Europe to Inner Asia. However, little is known about their economic foundations and the factors contributing to their extensive mobility.

Our results establish that sheep dairying accompanied the earliest forms of Eneolithic pastoralism in the North Caucasus. During the 4th millennium BCE, Maykop and early Yamnaya populations also focused dairying exclusively on sheep while reserving cattle for traction and other purposes. We observe a breakdown in livestock specialization and economic diversification of dairy herds coinciding with aridification during the subsequent late Yamnaya and North Caucasus Culture phases, followed by severe climate deterioration during the Catacomb and Lola periods. The need for additional pastures to support these herds may have driven the heightened mobility of the Middle and Late Bronze Age periods. Following a hiatus of more than 500 years, the North Caucasian steppe was repopulated by Iron Age societies with a broad mobile dairy economy. Our results confirm a subsistence strategy that included the dairying of sheep, goats, and cattle and the introduction of horse dairying.

**Author contributions:**

Christina Warinner, Wolfgang Haak, Sven Hansen, Emily Hammer, and Johannes Krause designed this study. Archaeological material and resources were provided by Sabine Reinhold, Sven Hansen, Alexey A. Kalmykov, Andrey Belinskiy, Alexandra Buzhilova, Natalia Berezina, Farhad Guliyev, Bertille Lyonnet, Parviz Gasimov, Tufan Axhundov, Bakhtiyar Jalilov, Jeyhun Eminli, Emil Iskandarov, Emily Hammer, Selin Nugent, Kerttu Majander, Päivi Onkamo, Kerkko Nordqvist, Natalia Shishlina, Elena Kaverzneva, Arkadiy I. Korolev, Aleksandr A. Khokhlov, Roman V. Smolyaninov, Rüdiger Krause, Svetlana V. Sharapova, Eliza Stolarzyk, Maria Karapetian

Ashley Scott processed dental calculus samples in the dedicated ancient protein clean lab at the Max Planck Institute for the Science of Human History in Jena, where they were

cataloged and weighed. Samples with a starting weight of > 5 mg were selected for analysis, and 5-13 mg were utilized for each protein extraction. In total, Ashley Scott processed 47 samples. Additionally, five samples included in this study were extracted by Richard Hagan. Extracted samples were sent to the Function Genomics Center in Zurich, Switzerland, where they were analyzed by tandem mass spectrometry. Ashley Scott analyzed the data, including quality control measures to authenticate results. Christina Warinner, Sabine Reinhold, Sven Hansen, Wolfgang Haak, and Taylor Hermes assisted Ashley Scott with interpreting the protein results in the larger context of the archaeological record. Ashley Scott wrote the main reports and initial manuscript with considerable contributions from Christina Warinner and Sabine Reinhold and minor contributions from Wolfgang Haak and Taylor Hermes, with input from all co-authors.

| | |
|---|---|
| Sample procurement, organization, and processing | 10% |
| Ancient protein laboratory work | 90% |
| Ancient protein bioinformatic processing | 100% |
| Ancient protein verification and quality control | 100% |
| Manuscript Construction and Organization | 75% |

Ashley Scott contributed 75% to the overall project, including organization, laboratory work, data analysis, and manuscript construction.

Note: Supplementary Data files can be accessed via the following DOI:
https://doi.org/10.1038/s41559-022-01701-6

Jena, 11.10.2021

________________________________

Ashley Scott

## 3.4. Manuscript D

**Exotic foods reveal contact between South Asia and the Near East during the 2nd millennium BCE**

Ashley Scott, Robert C. Power, Victoria Altmann-Wendling, Michal Artzy, Mario A. S. Martin, Stefanie Eisenmann, Richard Hagan, Domingo C. Salazar-García, Yossi Salmon, Dmitry Yegorov, Ianir Milevski, Israel Finkelstein, Philipp W. Stockhammer, Christina Warinner

**Published in *PNAS* (December 2020)**

**Synopsis**:
In Manuscript D, we applied proteomic and microremain techniques to the dental calculus of individuals buried in the Southern Levantine sites of Megiddo and Tel Erani; the first study of its kind in the ancient Near East. We confirmed the presence of staple crops such as wheat (*Triticum*), millet (Panicoideae), and date palm (*Phoenix)*. Additionally, we found robust evidence for sesame (*Sesamum*) proteins which confirmed that by the 2$^{nd}$ millennium BCE, sesame had become a staple oil-bearing crop in the Levant. We identified the earliest direct evidence for turmeric (*Curcuma*), which pushes back the earliest evidence of this spice in the region by ~900 years. We also report evidence for the consumption of soybean proteins. Additionally, we identified credible evidence for banana (*Musa*) in the Early Iron Age Levant, which provides a critical geographic link between the crop's origins in New Guinea (5$^{th}$ millennium BCE) and its appearance in West Africa (Cameroon) by the 1$^{st}$ millennium BCE

**Author contributions:**
Philipp Stockhammer designed the study with assistance from Ashley Scott, Robert Power, and Christina Warinner. The study consisted of both a microparticle and a proteomic component. Philipp Stockhammer Michal Artzy, and Yossi Salmon provided materials and resources, while Ashley Scott, Richard Hagan, Domingo Salazar-Garcia, and Stefanie Eisenmann contributed to data collection and organization. Ashley Scott performed the proteomic laboratory work at the dedicated ancient protein clean lab at the Max Planck Institute for the Science of Human History. Ashley Scott analyzed and authenticated the ancient protein results with feedback from Christina Warinner. Robert power analyzed samples for microparticle analysis with input from Christina Warinner. Philipp Stockhammer, Christina Warinner, Israel Finkelstein, Mario Martin, Ianir Milevski, Dmitry Yegorov, and Victoria Altmann-Wendling assisted Ashley Scott and Robert Power with data interpretation in the context of the archaeological record. Ashley Scott wrote the manuscript with significant contributions from Philipp Stockhammer, Christina Warinner, Robert Power, Victoria Altmann-Wendling, and assistance from all co-authors.

| | |
|---|---|
| Sample procurement, organization, and processing | 50% |
| Ancient protein laboratory work | 100% |
| Ancient protein bioinformatic processing | 100% |
| Ancient protein verification and quality control | 100% |
| Manuscript Construction and Organization | 75% |

Ashley Scott contributed 75% to the overall project, including organization, laboratory work, data analysis, and manuscript construction.

Note: Supplementary Data files can be accessed via the following DOI:
https://doi.org/10.1073/pnas.2014956117

Jena, 11.10.2021

________________________________

Ashley Scott

# 4. Manuscript A

**Manuscript Nr. : 1 (Manuscript A)**
**Title of the Manuscript:** Relief food subsistence revealed by microparticle and proteomic analyses of dental calculus from victims of the Great Irish Famine
**Authors:** Johnny Geber, Monica Tromp, Ashley Scott, Abigail Bouwman, Paolo Nanini, Jonas Grossman, Jessica Hendy, and Christina Warinner
**Bibliographic Information:** Geber J*, Tromp M*, Scott A*, Bouwman A, Nanni P, Grossman J, Hendy J, Warinner C. (2019) Relief Food Subsistence Revealed by Microparticle and Proteomics Analyses of Dental Calculus from Victims of the Great Irish Famine. *Proceedings of the National Academy of Sciences* USA 116(39):19380-19385.
**The candidate is:** ☐ First author, ☒ Co-first author, ☐ Corresponding author, ☐ coauthor
**Status:** Published
**Proportion (in %) of authors in the publication:** (to be indicated from 20%)

| Author | Concept | Data analysis | Experimental | Manuscript composition | Material provision |
|---|---|---|---|---|---|
| Geber J | 20 % | | | 30% | 70% |
| Tromp M | 20% | 50% | 100% | 30% | |
| Scott A | 20% | 50% | | 30% | |
| Bouwman A | | | | | |
| Nanni P | | | | | |
| Grossman J | | | | | |
| Hendy, J | 20% | | | | |
| Warinner C | 20% | | | | 20% |

# Relief food subsistence revealed by microparticle and proteomic analyses of dental calculus from victims of the Great Irish Famine

Jonny Geber[a,1,2], Monica Tromp[b,c,1], Ashley Scott[d,1], Abigail Bouwman[e], Paolo Nanni[f], Jonas Grossmann[f], Jessica Hendy[d,g], and Christina Warinner[d,h]

[a]School of History, Classics & Archaeology, University of Edinburgh, Edinburgh EH8 9AG, United Kingdom; [b]Department of Archaeology, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; [c]Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin 9016, New Zealand; [d]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany; [e]Institute of Evolutionary Medicine, University of Zürich, 8057 Zürich, Switzerland; [f]Functional Genomics Centre Zurich, University of Zürich/Eidgenössische Technische Hochschule Zürich, 8057 Zürich, Switzerland; [g]Department of Archaeology, University of York, York YO1 7EP, United Kingdom; and [h]Department of Anthropology, Harvard University, Cambridge, MA 02138

Edited by Dolores R. Piperno, Smithsonian Institution, Washington, DC, and approved August 12, 2019 (received for review May 22, 2019)

**Food and diet were class markers in 19th-century Ireland, which became evident as nearly 1 million people, primarily the poor and destitute, died as a consequence of the notorious Great Famine of 1845 to 1852. Famine took hold after a blight *(Phytophthora infestans)* destroyed virtually the only means of subsistence—the potato crop—for a significant proportion of the population. This study seeks to elucidate the variability of diet in mid–19th-century Ireland through microparticle and proteomic analysis of human dental calculus samples ($n$ = 42) from victims of the famine. The samples derive from remains of people who died between August 1847 and March 1851 while receiving poor relief as inmates in the union workhouse in the city of Kilkenny (52°39′ N, −7°15′ W). The results corroborate the historical accounts of food provisions before and during the famine, with evidence of corn (maize), potato, and cereal starch granules from the microparticle analysis and milk protein from the proteomic analysis. Unexpectedly, there is also evidence of egg protein—a food source generally reserved only for export and the better-off social classes—which highlights the variability of the prefamine experience for those who died. Through historical contextualization, this study shows how the notoriously monotonous potato diet of the poor was opportunistically supplemented by other foodstuffs. While the Great Irish Famine was one of the worst subsistence crises in history, it was foremost a social disaster induced by the lack of access to food and not the lack of food availability.**

bioarchaeology | paleoethnobotany | microfossil | potato | poverty

Ireland in the 19th century was characterized by political turbulence, economic decline, and drastic social and demographic changes (1). Poverty and destitution were widespread, and a highly stratified society denoted social class from a variety of factors, including diet. The potato (*Solanum tuberosum*), especially the Irish lumper cultivar, was a staple for the poor and destitute (2). For numerous reasons, by the 1840s, ∼40% of the population of Ireland had become utterly dependent on the potato for subsistence. The dependency on a single food source exposed people to significant risks and culminated in a devastating famine between 1845 and 1852 when a fungal blight (*Phytophthora infestans*) caused widespread destruction of the potato crop (Fig. 1). Nearly 1 million people died as a direct consequence of famine-induced starvation and disease. An estimated 200,000 of these famine-related deaths took place in the union workhouses (3), which were institutions for poverty relief that were introduced to Ireland following the Irish Poor Law Act of 1838 (4). When the potato harvest failed, vast quantities of food were imported as poor relief (5). This food was primarily made up of oatmeal from Britain and so-called "Indian meal" (maize) from the United States. The latter was to become an even more distinct social marker of poverty than the potato.

While the historical records of Victorian-period Ireland are vast, they generally do not reflect on the variability of the laboring classes' diet on an individual level. The aim and objectives of this study are to elucidate aspects of food availability and consumption patterns among the socially marginalized and impoverished in mid–19th-century Ireland through microparticle and proteomic analyses of dental calculus. The calculus derives from a subsample ($n$ = 42) of archaeological skeletons of individuals ($n$ = 970) who died while receiving poor relief in the union workhouse in Kilkenny City (*SI Appendix*, Fig. S1) between August 1847 and March 1851. The precise date of the remains—and the abundant historical resources pertaining to this period—enables a unique opportunity to use dietary analysis of dental calculus to evaluate both the historical and bioarchaeological sources and provide a holistic view of the condition of the mid–19th-century poor and destitute.

## Results

**Microparticles.** Microparticle analysis was conducted on all samples, of which 41 contained a total of 383 starch granules (≥1 granule

### Significance

**This study provides direct evidence of the dependency on relief food in Ireland around the time of the Great Famine (1845 to 1852) through dental calculus analysis of archaeological human remains. The findings show a dominance of corn (maize) and milk from the identified foodstuffs and corroborate the contemporaneous historical accounts of diet and subsistence. It shows that microparticle and proteomic analyses, even when based on small archaeological samples, can provide a valid snapshot of dietary patterns and food consumption. The occurrence of egg protein, generally only included in the diet for the better-off social classes, also highlights how these analytical techniques can provide unanticipated insights into the variability of diet in historical populations.**

Author contributions: J. Geber and M.T. designed research; J. Geber, M.T., A.S., P.N., and J.H. performed research; A.B. and C.W. contributed new reagents/analytic tools; J. Geber, M.T., A.S., J. Grossmann, and C.W. analyzed data; and J. Geber, M.T., A.S., J.H., and C.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository, http://www.ebi.ac.uk/pride (dataset identifier PXD015002).

[1]J. Geber, M.T., and A.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: jonny.geber@ed.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1908839116/-/DCSupplemental.

PNAS PNAS PNAS PNAS

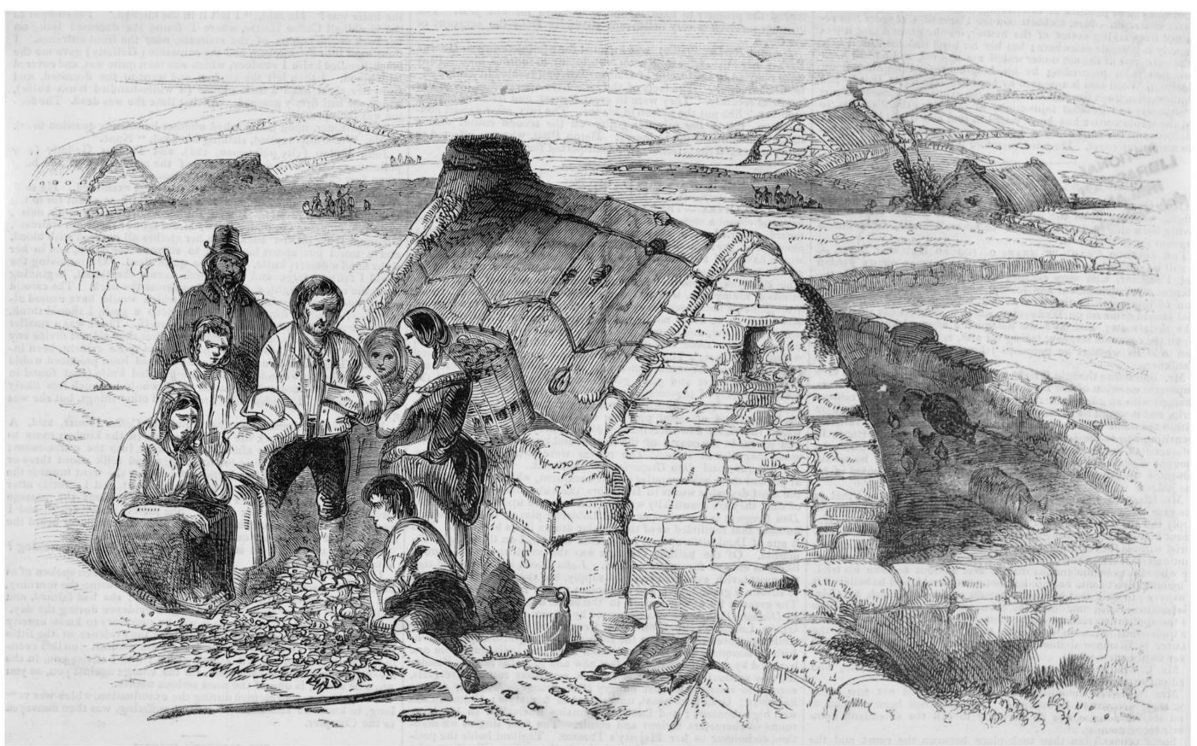


**Fig. 1.** "Destitution in Ireland. Failure of the potato crop" illustration was published in *The Pictorial Times* on 22 August 1846 (52). Image courtesy of the National Library of Ireland.

per sample); most of them were identified to 6 probable taxa (Table 1). Any granules measuring less than 10 μm were classified as unknown due to the difficulty in clearly distinguishing diagnostic features at this scale using light microscopy. The most dominant plant species identified was likely corn (*Zea mays*), which amounted to a total of 164 starch granules present in 38 samples (92.7%). There were 41 starch granules identified as probable oat (*Avena* spp.) present in 22 samples (53.7%). Probable wheat (*Triticum* spp.) granules ($n = 21$) were identified in 16 samples (39.0%). Five granules were identified as potato (*S. tuberosum*), and 5 were identified as probable malted barley (*Hordeum* spp.) (Fig. 2).

In addition to the identified taxa described above, there were 7 granules in 6 samples (14.6%) that derive from unidentified cereals (Poaceae), and there were a total of 140 granules across 37 samples (90.2%) that are of unknown origin, are less than 10 μm, or displayed nondiagnostic features. Of these unidentified granules, the majority are likely to derive from Poaceae starch. Also included among these are damaged starch microparticles (in all cases, they were not damaged to the point of losing visible extinction crosses). While it is well known that heat damages starch granules, studies have shown that it is still possible for some granules to survive unless the foods were boiled for more than 60 min (6–8), and these damaged granules may show the effects of cooking and food processing (6, 7).

Other findings include fungal material, hair, fibers, insect parts, and other unidentified plant and organic material (Dataset S1).

**Table 1. Summary table of number of starch granules observed from dental calculus samples ($n$ = 42) from the Kilkenny Union Workhouse by sex and age groups**

| | Sex | | | Age group (y) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Taxa | M ($n$ = 22) | F ($n$ = 19) | ? ($n$ = 1) | 13–17 ($n$ = 2) | 18–25 ($n$ = 4) | 26–35 ($n$ = 12) | 36–45 ($n$ = 18) | ≥46 ($n$ = 6) | Total ($n$ = 42) |
| Corn (*Z. mays*) | 95 | 66 | 3 | 6 | 17 | 44 | 73 | 24 | 164 |
| Oat (*Avena* spp.) | 20 | 21 | 0 | 2 | 8 | 7 | 13 | 11 | 41 |
| Wheat (*Triticum* spp.) | 9 | 12 | 0 | 0 | 5 | 4 | 10 | 2 | 21 |
| Malted barley (*Hordeum* spp.) | 3 | 2 | 0 | 1 | 0 | 1 | 3 | 0 | 5 |
| Cereal (Poaceae) | 4 | 2 | 1 | 1 | 0 | 0 | 3 | 3 | 7 |
| Potato (*S. tuberosum*) | 3 | 2 | 0 | 0 | 1 | 1 | 2 | 1 | 5 |
| Indeterminable | 88 | 50 | 2 | 14 | 10 | 29 | 59 | 28 | 140 |
| Total | 222 | 155 | 6 | 24 | 41 | 86 | 163 | 69 | 383 |

F, female; M, male; ?, indeterminable.

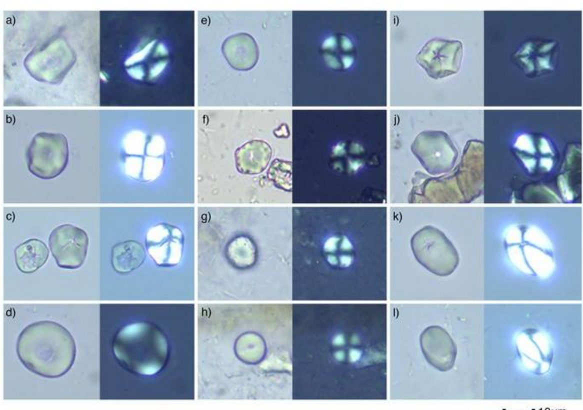



**Fig. 2.** Examples of starches observed in the microparticle analysis: (*A–C*) corn (*Z. mays*), (*D* and *E*) malted barley (*Hordeum* spp.), (*F–H*) wheat (*Triticum* spp.), (*I* and *J*) oat (*Avena* spp.), and (*K* and *L*) potato (*S. tuberosum*).

Some of the unidentified organic material may be chaff from unprocessed or poorly processed cereals, and it possibly reflects food adulteration. It may also be microparticles that derive from nondietary sources (9), such as from beddings that, for the poor, were composed of nothing more than chaff and straw both before and during institutionalization (10). The hair and fibers were present in 9 samples, with a notably high concentration in dental calculus from a 26- to 35-y-old female (CCCXXI).

**Proteomics.** Dental calculus was extracted from 14 individuals, and they were selected based on having deposits of dental calculus ≥1 mg for potential analysis. Seven samples contained animal (milk or egg) proteins (*SI Appendix*, Fig. S4), and 6 samples contained glyceraldehyde 3-phosphate dehydrogenase (GAPDH), a ubiquitous cellular protein of probable plant origin (Table 2) (11). Modifications expected for ancient/historical samples, such as deamidation (N, Q), were observed (Dataset S2). Peptides from the milk whey protein beta-lactoglobulin (BLG) were positively identified in 7 of the samples originating from 5 males and 2 females. Taxonomically diagnostic sequence variants in the BLG peptides confirmed the specific consumption of cattle (*Bos*; $n = 5$) and goat (*Capra*; $n = 1$) milk, and 3 individuals consumed bovid milk of ambiguous origin (cattle/sheep or sheep/goat). Cattle milk seems to have been the dominant milk source, and it is likely that all 7 individuals consumed cattle milk. Peptides from the egg protein ovalbumin, originating from domestic chicken (*Gallus*), were also identified in 2 males and 1 female, and each of these individuals also consumed milk. The cellular protein GAPDH has been previously found to be the most frequently recovered plant protein in archaeological dental calculus (12), and GAPDH peptides of probable plant origin were identified in 6 individuals (3 males, 3 females). Unfortunately, the peptide sequences are highly conserved among green plants (Viridiplantae) as well as a small number of bacterial species, and they are, therefore, not specific enough to identify individual taxa.

## Discussion

While the potato was the staple diet for the poor throughout Ireland, there were regional differences regarding agricultural and horticultural produce and dietary patterns across the island. In the west, where the social conditions were most deprived, oatmeal was rarely a dietary component, while the opposite was true in the northeast. Bread was mostly consumed in the east, and milk was a common component in the diet in the southwest, the southeast, and the north (13). Local agriculture in Kilkenny was dominated by grain and cattle production, one of the foundations of the relative prosperity of the city (14). Other than cereals and vegetables, markets were well supplied with sea fish, meat, and reportedly, even exotic fruits, such as peaches, melons, grapes, and pineapple (15). Much of the agricultural produce, however, was destined for the export market, and the diet of the average laborer in Kilkenny would corroborate what was seen elsewhere in Ireland. An account from the early 1840s stated that occupants on even sizable County Kilkenny farms of 70 to 100 acres would generally only have “potatoes and milk, stirabout, or oaten-bread and milk for breakfast and supper, and generally potatoes and butter, but once or twice herrings or butcher-meat, for dinner” (16) (stirabout is porridge made from oats, Indian meal, or both).

In accordance with the Poor Law, the composition of the Kilkenny Union Workhouse diet (*SI Appendix*, Table S1) was intentionally dire. Before 1846, potatoes were a staple and supplemented with milk rations that would have reflected the diet for many of the inmates before admission into the institution. During the famine, potatoes were chiefly replaced with Indian meal. In the

**Table 2. Summary table of the dietary proteins identified in dental calculus samples analyzed through proteomics ($n = 10$) by sex and age groups**

| Skeleton identification no. | Age group (y) | Sex | Extraction identification no. | Putative dietary protein | Total PSMs identified | Protein taxonomic assignment |
|---|---|---|---|---|---|---|
| CXCV | 36–45 | M | Z527 | BLG | 24 | Cattle (Bovinae) or sheep (*Ovis*) |
| CCXXVII | ≥46 | M | Z526 | BLG | 59 | Cattle (*Bos*) and goat (*Capra*) |
| CCXXVII | ≥46 | M | Z526 | GAPDH | 6 | Probable green plant (Viridiplantae) |
| CCXXVII | ≥46 | M | Z526 | Ovalbumin | 7 | Chicken (*Gallus*) |
| CCXXXIV | 26–35 | F | Z321 | GAPDH | 4 | Probable green plant (Viridiplantae) |
| CCLXXXII | 36–45 | M | Z523 | BLG | 10 | Cattle (*Bos*) |
| CCCXVIII | ≥46 | M | Z528 | GAPDH | 10 | Probable green plant (Viridiplantae) |
| CCCXCVIII | 36–45 | M | Z534 | BLG | 8 | Cattle (*Bos*) and sheep/goat (Caprinae) |
| DLXIV | 18–25 | F | Z533 | BLG | 3 | Cattle (*Bos*) |
| DLXIV | 18–25 | F | Z533 | GAPDH | 4 | Probable green plant (Viridiplantae) |
| DLXIV | 18–25 | F | Z533 | Ovalbumin | 4 | Chicken (*Gallus*) |
| DLXXII | 36–45 | F | Z529 | GAPDH | 10 | Probable green plant (Viridiplantae) |
| DCLXXXIII | 36–45 | M | Z532 | BLG | 11 | Cattle (*Bos*) |
| DCLXXXIII | 36–45 | M | Z532 | GAPDH | 7 | Probable green plant (Viridiplantae) |
| DCLXXXIII | 36–45 | M | Z532 | Ovalbumin | 6 | Chicken (*Gallus*) |
| DCCCXXIX | 26–35 | F | Z531 | BLG | 8 | Cattle (Bovinae) or sheep (*Ovis*) |

A peptide summary is provided in Dataset S2. F, female; M, male; PSM, peptide spectrum match.

latter years of the famine, a vegetable soup was introduced as part of the adult food rations, although at rations of only half a pint for dinner. This soup consisted of 5 ounces of rice, 5 ounces of oatmeal, 12 pounds of turnips (swedes), 2 pounds of parsnips, and onions with pepper and salt as required per gallon (17). Some meat was provided as part of the medical treatment in the workhouse infirmary (18), and for some periods, the minute books reveal the quantities of meat stored in the house. In the last week of February 1851, for instance, 390 pounds of meat were recorded in the food inventory. On those dates, the same records show that the union was providing indoor relief for 4,282 people! Other than flavor enhancers, such as pepper and allspice, the minute books also reveal other food products, such as arrowroot, sugar, broth, cocoa, port and sherry wine, whiskey, and porter. These are likely, just as the meat, to have been provisioned only for inmates treated in the workhouse infirmary.

**Historical Contextualization of Results.** The microparticle and proteomic analyses of the dental calculus samples exhibit a dominance of foodstuffs that are consistent with the 19th-century Irish laborer's diet and relief food consumed during the famine, suggesting that the calculus buildup in the analyzed subsample reflects tartar accumulations prior to and/or during life in the workhouse. This notion is primarily argued from the dominance of corn deriving from Indian meal, which was a central component of the workhouse diet during the famine, although it had also occasionally been distributed as poor relief in County Kilkenny on previous occasions (19). Stable isotope analyses conducted to date on a subsample of the skeletal remains have also indicated a C4 dietary input (maize), which further reflects the relief food dependency for the poor before and during the famine (20). The evidence of oat starch granules in the calculus samples could reflect both the relief food and/or the prefamine diet of these individuals.

The high occurrence of identifiable corn starch granules may also reflect the documented difficulty in processing the grain; Irish millers generally lacked the knowledge of how to sufficiently grind Indian corn, and the course grind may have allowed for the starch granules to survive the boiling process. The fact that the starch granules from corn dominate the samples also highlights how the experience of the famine was a prolonged period of suffering and struggle and that death was not instant. While Indian meal would have helped to combat hunger, it may ultimately also have caused further health deprivations. The corn was profoundly inferior in its nutritional value, as it, at the time, was improperly prepared, with a devastating effect on the health of the already weakened and frail. This was largely the result of a European rejection of adopting the traditional indigenous American method of preparing maize through the process of nixtamalization (alkaline cooking), which releases niacin from the grains (21). As a consequence, niacin deficiency—resulting in pellagra—is likely to have been an inadvertent consequence of the relief food distributed to the starving poor in Ireland during the famine (22).

As stated above, the workhouse diet included a soup, but its vegetable ingredients were not identified by protein or microscopic analysis in this study. This is likely due to the quick gelatinization process of vegetable starches (tubers and grains) during cooking (23, 24) as well as incomplete proteomic reference databases and the lack of identifiable microparticle production of some plants (onions, swedes). Starch granules from rice (*Oryza sativa*), another ingredient in the soup, could potentially be identified from microparticle analysis, as they gelatinize at much higher temperatures and have a higher lipid content (24). However, these granules would measure less than 10 μm in size (24) and therefore, would all be present among the unidentifiable microparticles found in this study. The absence of any grass/grain phytoliths in the dental calculus suggests that the plant foods that people ate were processed and not wild. Phytoliths are generally found in the husks and skin of grains, and in preindustrial archaeological dental calculus—where processing was minimal or done by hand—phytoliths are commonly found (25, 26).

The gelatinization of starch granules during cooking may also explain the low frequency of potato and the high frequency of corn starch granules in the calculus samples, although—naturally in the context of this study—this may also reflect crop availability. The gelatinization process of potato starch begins around a temperature of ~30 °C, and therefore, boiling, which was the preferred method for cooking potatoes in 19th-century Ireland, would have completely gelatinized the potato starch granules (27). Corn, however, has a much higher gelatinization temperature (28). The fact that some potato starch granules were nevertheless identified in a limited number of samples may reflect the way that the poor prepared the vegetable.

> The cabin-boiled potato was dressed in two ways: with and without the *bone* or the *moon*, as it is universally called by the genuine Irish. In the latter form, the potato was done to the heart, equally mealy throughout, and bursting its skin with fatness. This was the supper when children and young persons were to partake of the meal; but when much work was to be done, or a long fast to be endured, the heart or central nucleus of the potato was allowed, by checking the boil at a particular period, to remain parboiled, hard and waxy; and when the rest of the potato had been masticated in the usual manner, this hard lump, about the size of a small walnut, was bolted; and in this manner nearly a stone of the root was taken into the stomach of the Irish labourer per diem. (29)

The presence of the milk protein BLG in the calculus of half of the tested individuals supports the historical records of milk being an important liquid component of the poor laborer's diet. Somewhat surprising is that only 1 instance of BLG from goat (*Capra*) and a second from sheep/goat (Caprinae) were identified given that the goat—generally referred to as the "poor man's cow"—played a particularly important role in the Irish domestic economy due to, what was said, the "impoverished state of the country" (30). However, additional trypsin cut sites in caprine BLG may lead to smaller peptides that are less likely to be identified through liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis; thus, an absence of these peptides does not necessarily indicate that goat milk was not widely consumed. If indeed the specific identification of cattle milk as opposed to other ruminant taxa is accurate, the fact that milk from cattle is dominating the identifications may also suggest that it derives primarily from the workhouse diet, as milk was purchased by the Kilkenny Union at large quantities every week from a limited number of suppliers, and this is likely to indicate that it was derived from a larger and more commercial production setup.

**Diet Variability and Food Access during the Time of the Great Irish Famine.** The shame associated with poor relief in the form of food is mentioned in local folklore that tells of how receiving stirabout made from Indian meal was gravely humiliating for a great number of people who, prior to the blight, had considered themselves as "respectable farmers" (31). For many others, however, the desperate struggle for subsistence—and the associated perceived stigma—had been a constant reality of their existence long before the famine. A German visitor to Ireland gave one particularly evocative account from Kilkenny City where he arrived by coach in August 1837.

> At Kilkenny there is an old castle, and innumerable beggars. . .The coach is besieged by them, and their cries resounds from all sides, and in all gradations of old and young voices. . .I saw a mother pick up the gooseberry skins which one of the travellers had spit out, and put them into the mouth of her child. I never [before] saw any thing like this. (32)

For obvious reasons, the situation for the poor and destitute had declined gravely a decade later, and folklore tells of how people in Kilkenny, just as elsewhere in Ireland, had resorted to

eating grass and weeds during the famine (33). Local accounts also state that, while the poor and the laboring classes suffered greatly, the fertile lands in the county generated good crops of wheat during the years of the potato blight, but this food was bought up by the local gentry who made profits by exporting it for sale at English markets (34). Other recorded narratives, however, tell of how locally produced oats and wheat had, indeed, been distributed in County Kilkenny as famine relief food (35).

While major aspects of the relief food provided by the Kilkenny Union are evidenced through this study, there were also dietary components indicated from the microparticle and proteomic analyses that show aspects of the variability of the laborer's diet in 19th-century Ireland. Proteomic analysis identified the egg protein ovalbumin in the dental calculus of 3 individuals. Eggs were regarded as "luxury food" for the working classes (13) and thereby, also a dietary class marker. Although it was not uncommon for the peasantry to keep poultry, eggs were not eaten except for, perhaps, during celebratory occasions. Eggs were, just as much as the wheat crop, destined for the export market, or as one Irish laborer in the 1830s stated, "We eat none of them; they go to our rent, or to put a shoe on our foot, or a spade in our hands" (36).

The evidence of egg protein through this study is highlighting how the prefamine experience and living conditions might have varied for those people who were forced to resort to the Irish workhouses in the 1840s. The evidence of probable wheat in the dental calculus of some individuals from the Kilkenny Union Workhouse can also be viewed in the same light, although it is possible that the wheat derives from the occasional bread that was part of the workhouse diet or from an as yet unidentified local wild grass. Further references to the crop in famine-period records from Kilkenny tell of how wheat was being "crop-lifted" by desperate people (37), and there were also several reports of "bread riots" occurring in the city during the famine (38).

## Conclusion

The social, political, and economic historical background of the Great Irish Famine is very complex, and the term "The Irish Potato Famine" (generally never used in Ireland) is problematic for that reason. While the famine was a subsistence crisis initiated by the destruction of the potato crop blight, it was much more than a natural disaster relating to a single vegetable. Prevailing laissez-faire economic convictions profoundly influenced the political response to the food crisis that occurred, and vast amounts of food products were exported from Ireland while the poor starved (39). There was also an undoubtedly colonial dimension to these economic policies, where Ireland and its primarily Catholic population—which following the Act of Union in 1801, were governed directly by a British government from Westminster—were subjugated through various political means (40).

As made evident through this study, scientific evaluations using microparticle and proteomic analytical techniques of archaeological dental calculus can provide complimentary snapshots of overall diets even from exceptionally well-contextualized and well-recorded historical samples. This study suggests that diet variability of the poor was limited and expressed equally across both sexes among the poor in Ireland during the time of the Great Famine. When interpreting the findings in their historical and cultural context, the evidence of eggs (and probably also wheat) also indicates how the food subsistence pattern for the laboring classes was opportunistic. Contemporary 19th-century accounts would state that the potato was the food of choice for the poor (41), but this was a truth with modification. When commenting on whether the peasantry preferred potatoes over meal or bread, one laborer from the west of Ireland during the 1840s replied: "Why would we prefer that, that we feed our pigs on to better food? Don't you like it better yourself, and why shouldn't we? Never believe them that would want to make you think that we'd eat wet lumpers if we could get good bread" (42). This notion is further indicated from this study, which has provided nuanced insights into how socially marginalized people sought to maintain their subsistence while enduring a life of constant hardship and unimaginable struggles.

## Materials and Methods

A mass burial ground adjacent to the former Kilkenny Union Workhouse, Kilkenny City, Ireland, was excavated in 2006. A minimum of 970 individuals from 63 burial pits were recovered. The skeletal remains were analyzed following standard osteological methods. Dental calculus was sampled from individual teeth extracted from a subsample of 42 skeletons prior to the reburial in 2010 (*SI Appendix* has details).

Microparticles were extracted from 42 dental calculus samples (on average $\sim 3 \times 3$ mm) using the ethylenediaminetetraacetic acid (EDTA) extraction method described in Tromp et al. (43) and briefly reviewed in *SI Appendix*. Starch granules were analyzed primarily using published studies and descriptions (24, 44–50) as well as a reference collection; measurements and descriptions of relevant taxa are included in Dataset S3. Identifications were made based on the size (maximum width), shape, and distinguishing features, such as visible lamellae, surface pitting, pressure facets, and visible and/or cracked hilum (additional details are in *SI Appendix* and Datasets S3 and S5).

Proteins were extracted from decalcified dental calculus samples according to guidelines recommended by Hendy et al. (51). Extracted peptides were analyzed by LC-MS/MS using a Q-Exactive HF mass spectrometer (Thermo Scientific) coupled to an ACQUITY UPLC M-Class system (Waters AG) (*SI Appendix* has additional details). Tandem mass spectra were converted to Mascot generic files by MSConvert version 3.0.11781 with the 100 most intense peaks in each spectrum. All MS/MS samples were analyzed using Mascot (Matrix Science; version 2.6.0). Mascot was set up to search the databases SwissProt_2017_07.fasta and uniprot_trembl_2017_07, assuming the digestion enzyme trypsin and selecting the automatic decoy option (*SI Appendix* has additional details). Scaffold (version Scaffold_4.8.9; Proteome Software Inc.) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications at a local false discovery rate (FDR) of less than 1.0% (Scaffold Local FDR algorithm) and protein identifications at an FDR of less than 5.0% and supported by at least 3 identified peptides were accepted (*SI Appendix* has additional details). Raw data files are available through the ProteomeXchange Consortium via the PRIDE partner repository (accession no. PXD015002).

**ACKNOWLEDGMENTS.** We thank the editor and the anonymous reviewers for their constructive criticism, helpful comments, and suggestions, which have improved the paper. This study is made possible thanks to permission and support from the Irish Antiquities Division of the National Museum of Ireland (NMI). Thanks are also due to Dr. Regina Sexton and Dr. Laurence Geary for input on food culture in 19th-century Ireland, and to Dr. Christian Trachsel for assistance with protein mass spectrometry. The archaeological excavation at the Kilkenny Union Workhouse mass burials was undertaken by Margaret Gowen & Co. Ltd. under license 05E0435 from the NMI. Laboratory analyses undertaken for this study were funded by the Max Planck Institute for the Science of Human History; the Department of Anatomy of the University of Otago; research grants from Johan and Jakob Söderberg's Foundation and the Royal Irish Academy in partnership with the National Monuments Service; Wellcome Trust Grant 096435/Z/11/Z; and US NSF Grant BCS-1523264.

1. C. Ó Gráda, *Ireland before and after the Famine: Explorations in Economic History, 1800–1925* (Manchester University Press, 1993).
2. P. M. A. Bourke, The use of the potato crop in pre-Famine Ireland. *J. Stat. Soc. Inq. Soc. Irel.* **21**, 72–96 (1968).
3. C. Ó Gráda, "Mortality and the Great Famine" in *Atlas of the Great Irish Famine, 1845–52*, J. Crowley, W. J. Smyth, M. Murphy, Eds. (Cork University Press, 2012), pp. 170–179.
4. J. O'Connor, *The Workhouses of Ireland: The Fate of Ireland's Poor* (Anvil Books, 1995).
5. M. S. Harzallah, Food supply and economic ideology: Indian corn relief during the second year of the Great Irish Famine (1847). *Historian* **68**, 305–323 (2006).
6. A. G. Henry, H. F. Hudson, D. R. Piperno, Changes in starch grain morphologies from cooking. *J. Archaeol. Sci.* **36**, 915–922 (2009).
7. A. Crowther, The differential survival of native starch during cooking and implications for archaeological analyses: A review. *Archaeol. Anthropol. Sci.* **4**, 221–235 (2012).
8. S. L. Schnorr, A. N. Crittenden, A. G. Henry, Impact of brief roasting on starch gelatinization in whole foods and implications for plant food nutritional ecology in human evolution. *Ethnoarchaeology* **8**, 30–56 (2016).
9. A. Radini, E. Nikita, S. Buckley, L. Copeland, K. Hardy, Beyond food: The multiple pathways for inclusion of materials into ancient dental calculus. *Am. J. Phys. Anthropol.* **162** (suppl. 63), 71–83 (2017).

PNAS PNAS PNAS PNAS


10. Poor Law Commission, Poor Inquiry (Ireland). Appendix (E.) (Majesty's Stationary Office, 1836).
11. A. Scott, C. Warinner. Project PXD015003. Proteomics Identifications (PRIDE) Archive. http://doi.org/10.6019/PXD015002. Deposited 13 August 2019.
12. J. Hendy *et al*., Proteomic evidence of dietary sources in ancient dental calculus. *Proc. Biol. Sci.* **285**, 20180977 (2018).
13. L. Kennedy, P. S. Ell, E. M. Crawford, L. A. Clarkson, *Mapping the Great Irish Famine: A Survey of the Famine Decades* (Four Courts Press, 1999).
14. D, On the agriculture of the county of Kilkenny. *Q. J. Agric.* **9**, 475–495 (1839).
15. W. Tighe, *Statistical Observations Relative to the County of Kilkenny, Made in the Years 1800 & 1801* (Graisberry and Campbell, 1802).
16. Anonymous, *The Parliamentary Gazetteer of Ireland, Adapted to the New Poor-Law, Franchise, Municipal and Ecclesiastical Arrangements, and Compiled with a Special Reference to the Lines of Railroad and Canal Communication, as Existing in 1844–45; Illustrated by a Series of Maps, and Other Plates; and Presenting the Results, in Detail, of the Census of 1841, Compared with That of 1831* (A. Fullarton and Co., 1846), vol. 2.
17. Board of Guardians, *The Kilkenny and Leinster Commercial and Literary Advertiser*, 13 December 1848.
18. Kilkenny Union Board of Guardians, "Minutes of Kilkenny Union Board of Guardians meeting 5 October" (Kilkenny County Library, Local Studies, 1842).
19. T. de Bhaldraithe, *The Diary of an Irish Countryman 1827–1835. A Translation of Cín Lae Amhlaoibh* (Mercier Press, 1979).
20. J. Beaumont, J. Montgomery, The Great Irish Famine: Identifying starvation in the tissues of victims using stable isotope analysis of bone and incremental dentine collagen. *PLoS One* **11**, e0160065 (2016).
21. E. C. Ellwood, M. P. Scott, W. D. Lipe, R. G. Matson, J. G. Jones, Stone-boiling maize with limestone: Experimental results and implications for nutrition among SE Utah preceramic groups. *J. Archaeol. Sci.* **40**, 35–44 (2013).
22. E. M. Crawford, "Indian meal and pellagra in nineteenth-century Ireland" in *Irish Population, Economy, and Society*, J. M. Goldstrom, L. A. Clarkson, Eds. (Essays in Honour of the Late K. H. Connell, Clarendon Press, 1981), pp. 113–133.
23. R. F. Tester, J. Karkalas, Swelling and gelatinization of oat starches. *Cereal Chem.* **73**, 271–277 (1996).
24. J. Waterschoot, S. V. Gomand, E. Fierens, J. A. Delcour, Production, structure, physicochemical and functional properties of maize, cassava, wheat, potato and rice starches. *Starke* **67**, 14–29 (2015).
25. E. Norström, R. Gustavsson, F. Molin, S. Gummesson, Micro-fossil analysis of Mesolithic human dental calculus, Motala, Sweden–Indications of health status and paleo-diet. *J. Archaeol. Sci. Rep.* **67**, 101866 (2019).
26. A. Radini, E. Nikita, L. M. Shillito, "Human dental calculus and a medieval urban environment" in *Objects, Environment, and Everyday Life in Medieval Europe*, B. Jervis, L. G. Broderick, I. G. Sologestoa, Eds. (Brepols Publishers, 2016), pp. 297–313.
27. T. Shiotsubo, Gelatinization temperature of potato starch at the equilibrium state. *Agric. Biol. Chem.* **48**, 1–7 (1984).
28. M. Schirmer, M. Jekle, T. Becker, Starch gelatinization and its complexity for analysis. *Starke* **67**, 30–41 (2015).
29. W. Wilde, The food of the Irish. Dublin Univ. Mag. 43, 127–146 (1854).
30. J. Wilson, *British Farming: A Description of the Mixed Husbandry of Great Britain* (Adam and Charles Black, 1862).
31. Irish Folklore Commission/Department of Education and the National Irish Teacher's Organisation, "'Famine times,' told by an unknown informant, Paulstown, Co. Kilkenny" in *Schools' Folklore Collection* (National Folklore Collection, University College Dublin, 1937–1938), vol. 0861, p. 396.
32. F. von Raumer, *England in 1835: Being a Series of Letters Written to Friends in Germany, during a Residence in London and Excursions into the Provinces* (Carey, Lea, and Blanchard) 1836).
33. Irish Folklore Commission/Department of Education and the National Irish Teacher's Organisation, "'The Famine,' told by Mr Thomas Cuddihy (50 years), Pilltown, Co. Kilkenny to Mr John Ryan (12 years), Pilltown, Co. Kilkenny" in *Schools' Folklore Collection* (National Folklore Collection, University College Dublin, 1937–1938), vol. 0842, p. 259.
34. Irish Folklore Commission/Department of Education and the National Irish Teacher's Organisation, "'Knocktopher,' told by Fr Dwyer, Knocktopher, Co. Kilkenny to L. Geoghegan, Knocktopher, Co. Kilkenny" in *Schools' Folklore Collection* (National Folklore Collection, University College Dublin, 1937–1938), vol. 0848, pp. 304–322.
35. Irish Folklore Commission/Department of Education and the National Irish Teacher's Organisation, "Historic–Famine period–Black '47,' told by Mr. Denis Carroll (60 years), Templemartin, Co. Kilkenny" in *Schools' Folklore Collection* (National Folklore Collection, University College Dublin, 1937–1938), vol. 0859, pp. 373–374.
36. C. I. Johnstone, "True Tales of the Irish Peasantry, as Related by Themselves" Selected from the Report of the Poor Law Commissioners (William Tait, 1836).
37. Anonymous, News of the week. Ireland. Commencement of crop-lifting. *The Economist, Weekly Commercial Times, Bankers' Gazette, and Railway Monitor: A Political, Literary, and General Newspaper* **8**, 937–938 (1850).
38. Food riots in Kilkenny. *Morning Advertiser*, 12 June 1848.
39. C. Kinealy, Food exports from Ireland 1846–47. *Hist. Irel.* **5**, 32–36 (1997).
40. T. McDonough, Ed., *Was Ireland a Colony? Economics, Politics, and Culture in Nineteenth-Century Ireland* (Irish Academic Press, 2005).
41. A. Bourke, *'The Visitation of God'? The Potato and the Great Irish Famine*, J. Hill, C. Ó. Gráda, Eds. (Lilliput Press, 1993).
42. T. M. Ray, *Report on the Irish Coercion Bill, the Causes of Discontent in Ireland, Condition of the People, Comparative Criminality with England, Remedial Measures, Etc. Etc* (J. Browne, 1846).
43. M. Tromp, H. Buckley, J. Geber, E. Matisoo-Smith, EDTA decalcification of dental calculus as an alternative means of microparticle extraction from archaeological samples. *J. Archaeol. Sci. Rep.* **14**, 461–466 (2017).
44. R. Torrence, H. Barton, Eds., *Ancient Starch Research* (Left Coast Press, 2006).
45. A. G. Henry, D. R. Piperno, Using plant microfossils from dental calculus to recover human diet: A case study from Tell al-Raqā'I, Syria. *J. Archaeol. Sci.* **35**, 1943–1950 (2008).
46. T. Juhola, H. Etu-Sihvola, T. Näreoja, J. Ruohonen, Starch analysis reveals starchy foods and food processing from Finnish archaeological artefacts. *Fennosc. Archaeol.* **31**, 79–100 (2014).
47. D. Nadel, D. R. Piperno, I. Holst, A. Snir, E. Weiss, New evidence for the processing of wild ceral grains at Ohalo II, a 23 000-year-old campsite on the shore of the Sea of Galilee, Israel. *Antiquity* **86**, 990–1003 (2012).
48. D. R. Piperno, E. Weiss, I. Holst, D. Nadel, Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature* **430**, 670–673 (2004).
49. X. Yang, L. Perry, Identification of ancient starch grains from the tribe Triticeae in the North China Plain. *J. Archaeol. Sci.* **40**, 3170–3177 (2013).
50. M. Li, X. Yang, Q. Ge, X. Ren, Z. Wan, Starch grain analysis of stone knives from Changning site, Qinghai Province, Northwest China. *J. Archaeol. Sci.* **40**, 1667–1672 (2013).
51. J. Hendy *et al*., A guide to ancient protein studies. *Nat. Ecol. Evol.* **2**, 791–799 (2018).
52. Irish agitation and destitution. *The Pictorial Times*, 22 August 1846.

# 5. Manuscript B

**Manuscript Nr. : 2 (Manuscript B)**

**Title of the Manuscript:** The genomic origins of the Bronze Age Tarim Basin mummies

**Authors:** Fan Zhang*, Chao Ning*, Ashley Scott*, Qiaomei Fu, Rasmus Bjørn, Wenying Li, Dong Wei, Wenjun Wang, Linyuan Fan, Idilisi Abuduresule, Xingjun Hu, Qiurong Ruan, Alipujiang Niyazi, Guanghui Dong, Peng Cao, Feng Liu, Qingyan Dai, Xiaotian Feng, Ruowei Yang, Zihua Tang, Pengcheng Ma, Chunxiang Li, Shizhu Gao, Yang Xu, Sihao Wu, Shaoqing Wen, Hong Zhu, Hui Zhou, Martine Robeets, Vikas Kumar, Johannes Krause, Christina Warinner, Choongwon Jeong, Yinqiu Cui

**Bibliographic Information:** Zhang F, Ning C, Scott A, Fu Q, Bjorn R, et al. (2021) The genomic origins of the Bronze Age Tarim Basin mummies. *Nature.* [in press]

**The candidate is:** ☐ First author, ☒ Co-first author, ☐ Corresponding author, ☐ coauthor

**Status:** Accepted, in press

**Proportion (in %) of authors in the publication:** (to be indicated from 20%)

| Author | Concept | Data analysis | Experimental | Manuscript composition | Material provision |
|---|---|---|---|---|---|
| Zhang F | | | | | |
| Ning C | | 40% DNA | | 20% | |
| **Scott A** | | 100% (protein) | 100% (proteins) | 20% | |
| Fu Q | | | | | 30% |
| Bjørn R | | | | | |
| Li W | | | | | |
| Wei D | | | | | |
| Wang W | | | | | |
| Fan L | | | | | |
| Abuduresule I | | | | | |
| Hu X | | | | | |
| Ruan Q | | | | | |
| Niyazi A | | | | | |
| Dong G | | | | | |
| Cao P | | | | | |
| Liu F | | | | | |
| Dai Q | | | | | |
| Feng X | | | | | |
| Yang R | | | | | |
| Tang Z | | | | | |
| Ma P | | | | | |
| Li C | | | | | |
| Gao S | | | | | |
| Xu Y | | | | | |
| Wu S | | | | | |
| Wen S | | | | | |
| Zhu H | | | | | |
| Zhou H | | | | | |
| Robeets M | | | | | |
| Kumar V | | | | | |
| Krause J | 20% | | | | |
| Warinner C | 20% | | | 20% | |
| Jeong C | 20% | 40% (DNA) | | 20% | |
| Cui Y | 20% | | | | 40% |

Article

# The genomic origins of the Bronze Age Tarim Basin mummies

https://doi.org/10.1038/s41586-021-04052-7

Received: 19 November 2020

Accepted: 23 September 2021

Published online: 27 October 2021

Open access

Check for updates

Fan Zhang[1,15], Chao Ning[2,15] ✉, Ashley Scott[2,15], Qiaomei Fu[3], Rasmus Bjørn[2], Wenying Li[4], Dong Wei[5], Wenjun Wang[3], Linyuan Fan[1], Idilisi Abuduresule[4], Xingjun Hu[4], Qiurong Ruan[4], Alipujiang Niyazi[4], Guanghui Dong[6], Peng Cao[3], Feng Liu[3], Qingyan Dai[3], Xiaotian Feng[3], Ruowei Yang[3], Zihua Tang[7], Pengcheng Ma[1], Chunxiang Li[1], Shizhu Gao[8], Yang Xu[1], Sihao Wu[1], Shaoqing Wen[9], Hong Zhu[5], Hui Zhou[1], Martine Robbeets[2], Vikas Kumar[3], Johannes Krause[2,10] ✉, Christina Warinner[2,11] ✉, Choongwon Jeong[12] ✉ & Yinqiu Cui[1,13,14] ✉

The identity of the earliest inhabitants of Xinjiang, in the heart of Inner Asia, and the languages that they spoke have long been debated and remain contentious[1]. Here we present genomic data from 5 individuals dating to around 3000–2800 bc from the Dzungarian Basin and 13 individuals dating to around 2100–1700 bc from the Tarim Basin, representing the earliest yet discovered human remains from North and South Xinjiang, respectively. We find that the Early Bronze Age Dzungarian individuals exhibit a predominantly Afanasievo ancestry with an additional local contribution, and the Early–Middle Bronze Age Tarim individuals contain only a local ancestry. The Tarim individuals from the site of Xiaohe further exhibit strong evidence of milk proteins in their dental calculus, indicating a reliance on dairy pastoralism at the site since its founding. Our results do not support previous hypotheses for the origin of the Tarim mummies, who were argued to be Proto-Tocharian-speaking pastoralists descended from the Afanasievo[1,2] or to have originated among the Bactria–Margiana Archaeological Complex[3] or Inner Asian Mountain Corridor cultures[4]. Instead, although Tocharian may have been plausibly introduced to the Dzungarian Basin by Afanasievo migrants during the Early Bronze Age, we find that the earliest Tarim Basin cultures appear to have arisen from a genetically isolated local population that adopted neighbouring pastoralist and agriculturalist practices, which allowed them to settle and thrive along the shifting riverine oases of the Taklamakan Desert.

As part of the Silk Road and located at the geographic confluence of Eastern and Western cultures, the Xinjiang Uyghur Autonomous Region (henceforth Xinjiang) has long served as a major crossroads for trans-Eurasian exchanges of people, cultures, agriculture and languages[1,5–9]. Bisected by the Tianshan mountains, Xinjiang can be divided into two subregions referred to as North Xinjiang, which contains the Dzungarian Basin, and South Xinjiang, which contains the Tarim Basin (Fig. 1). The Dzungarian Basin in the north consists of the Gurbantünggüt Desert, which is surrounded by a vast expanse of grasslands traditionally inhabited by mobile pastoralists. The southern part of Xinjiang consists of the Tarim Basin, a dry inland sea that now forms the Taklamakan Desert. Although mostly uninhabitable, the Tarim Basin also contains small oases and riverine corridors, fed by runoff from thawing glacier ice and snow from the surrounding high mountains[4,10,11].

Within and around the Dzungarian Basin, pastoralist Early Bronze Age (EBA) Afanasievo (3000–2600 bc) and Chemurchek (or Qiemu'erqieke) (2500–1700 bc)[12] sites have been plausibly linked to the Afanasievo herders of the Altai–Sayan region in southern Siberia (3150–2750 bc), who in turn have close genetic ties with the Yamnaya (3500–2500 bc) of the Pontic–Caspian steppe located 3,000 km to the west[13–15]. Linguists have hypothesized that the Afanasievo dispersal brought the now extinct Tocharian branch of the Indo-European language family eastwards, separating it from other Indo-European languages by the third or fourth millennium bc (ref. [16]). However, although Afanasievo-related ancestry has been confirmed among Iron Age Dzungarian populations (around 200–400 bc)[7], and Tocharian is recorded in Buddhist texts from the Tarim Basin dating to ad 500–1000 (ref. [13]), little is known about earlier Xinjiang populations and their possible genetic relationships with the Afanasievo or other groups.

[1]School of Life Sciences, Jilin University, Changchun, China. [2]Max Planck Institute for the Science of Human History, Jena, Germany. [3]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, China. [4]Xinjiang Institute of Cultural Relics and Archaeology, Ürümqi, China. [5]School of Archaeology, Jilin University, Changchun, China. [6]MOE Key Laboratory of Western China's Environmental Systems, College of Earth & Environmental Sciences, Lanzhou University, Lanzhou, China. [7]Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China. [8]College of Pharmacia Sciences, Jilin University, Changchun, China. [9]Institute of Archaeological Science, Fudan University, Shanghai, China. [10]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [11]Department of Anthropology, Harvard University, Cambridge, MA, USA. [12]School of Biological Sciences, Seoul National University, Seoul, Republic of Korea. [13]Key Laboratory for Evolution of Past Life and Environment in Northeast Asia, Ministry of Education, Jilin University, Changchun, China. [14]Research Center for Chinese Frontier Archaeology of Jilin University, Jilin University, Changchun, China. [15]These authors contributed equally: Fan Zhang, Chao Ning, Ashley Scott. ✉e-mail: ning@shh.mpg.de; krause@shh.mpg.de; warinner@shh.mpg.de; cwjeong@snu.ac.kr; cuiyq@jlu.edu.cn

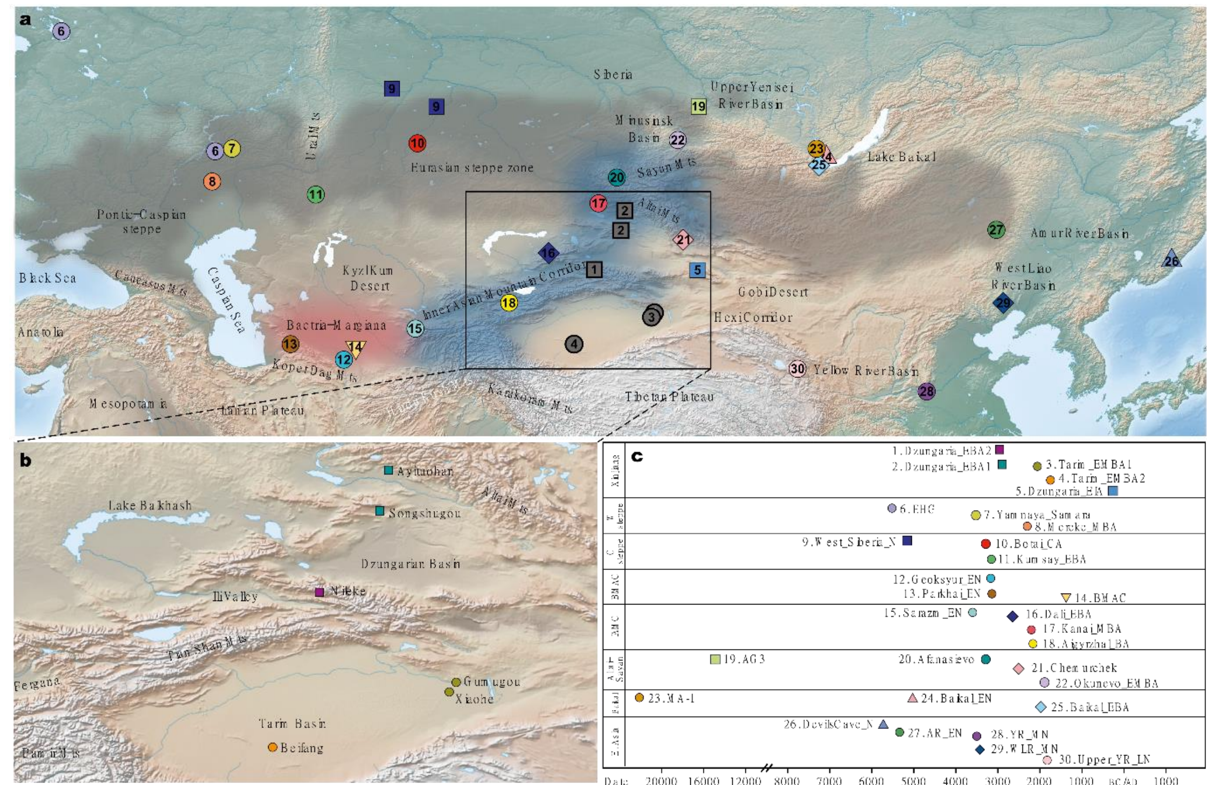



**Fig. 1 | Overview of the Xinjiang Bronze Age archaeological sites analysed in this study. a**, Overview of key Eurasian geographic regions, features and archaeological sites discussed in the text; new sites analysed in this study are shown in grey. **b**, Enhanced view of Xinjiang and the six new sites analysed in this study. **c**, Timeline of the sites in **a**. The timeline is organized by region, and the median date for each studied group is shown. The basemaps in **a** and **b** were obtained from the Natural Earth public domain map dataset (https://www.naturalearthdata.com/downloads/10m-raster-data/10m-cross-blend-hypso/). In the group labels, the suffixes represent the archaeological time periods of each group: N, Neolithic; EN, MN and LN, Early, Middle and Late Neolithic, respectively; EN, Eneolithic for Geoksyur, Parkhai and Sarazm; CA, Chalcolithic Age; BA, Bronze Age; MBA, Middle Bronze Age; EIA, Early Iron Age. MA-1, Mal'ta; EHG, Eastern European hunter-gatherers.

Since the late 1990s, the discovery of hundreds of naturally mummified human remains dating to around 2000 bc to ad 200 in the Tarim Basin has attracted international attention due to their so-called Western physical appearance, their felted and woven woollen clothing, and their agropastoral economy that included cattle, sheep/goats, wheat, barley, millet and even kefir cheese[16–19]. Such mummies have now been found throughout the Tarim Basin, among which the earliest are those found in the lowest layers of the cemeteries at Gumugou (2135–1939 bc), Xiaohe (1884–1736 bc) and Beifang (1785–1664 bc) (Fig. 1, Extended Data Fig. 1 and Extended Data Table 1). These and related Bronze Age sites are grouped within the Xiaohe archaeological horizon on the basis of their shared material culture[13,16,20].

Multiple contrasting hypotheses have been suggested by scholars to explain the origins and Western elements of the Xiaohe horizon, including the Yamnaya/Afanasievo steppe hypothesis[16], the Bactrian oasis hypothesis[21] and the Inner Asian Mountain Corridor (IAMC) island biogeography hypothesis[4]. The Yamnaya/Afanasievo steppe hypothesis posits that the Afanasievo-related EBA populations in the Altai–Sayan mountains spread via the Dzungarian Basin into the Tarim Basin and subsequently founded the agropastoralist communities making up the Xiaohe horizon around 2000 bc (refs. [16,22,23]). By contrast, the Bactrian oasis hypothesis posits that the Tarim Basin was initially colonized by migrating farmers of the Bactria–Margiana Archaeological Complex (BMAC) (around 2300–1800 bc) from the desert oases of Afghanistan, Turkmenistan and Uzbekistan via the mountains of Central Asia. Support for this hypothesis is largely based on similarities in the agricultural and irrigation systems between the two regions that reflect adaptations to a desert environment, as well as evidence for the ritual use of *Ephedra* at both locations[3,21]. The IAMC island biogeography hypothesis similarly posits a mountain Central Asian origin for the Xiaohe founder population, but one linked to the transhumance of agropastoralists in the IAMC to the west and north of the Tarim Basin[4,24,25]. In contrast to these three migration models, the greater IAMC, which spans the Hindu Kush to Altai mountains, may have alternatively functioned as a geographic arena through which cultural ideas, rather than populations, primarily moved[25].

Recent archaeogenomic research has shown that Bronze Age Afanasievo of southern Siberia and IAMC/BMAC populations of Central Asia have distinguishable genetic profiles[15,26], and that these profiles are likewise also distinct from those of pre-agropastoralist hunter-gatherer populations in Inner Asia[2,5,27–30]. As such, an archaeogenomic investigation of Bronze Age Xinjiang populations presents a powerful approach for reconstructing the population histories of the Dzungarian and Tarim basins and the origins of the Bronze Age Xiaohe horizon. Examining the skeletal material of 33 Bronze Age individuals from sites in the Dzungarian (Nileke, Ayituohan and Songshugou) and Tarim (Xiaohe, Gumugou and Beifang) basins, we successfully retrieved ancient genome sequences from 5 EBA Dzungarian individuals (3000–2800 bc)

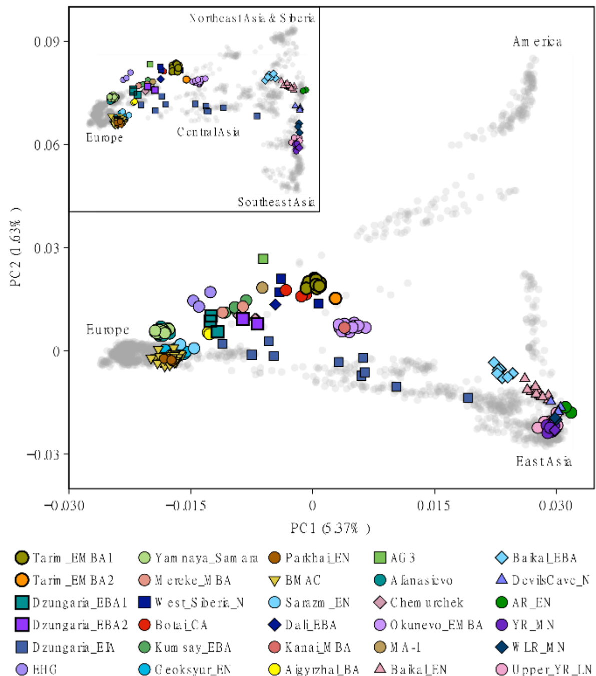



**Fig. 2 | Genetic structure of ancient and present-day populations included in this study.** Principal component analysis of ancient individuals projected onto Eurasian and Native American populations; the inset displays ancient individuals projected onto only Eurasian populations.

culturally assigned as Afanasievo, and genome-wide data from 13 Early–Middle Bronze Age (EMBA) Tarim individuals (2100–1700 bc) belonging to the Xiaohe horizon (Extended Data Table 1 and Supplementary Data 1A). We additionally report dental calculus proteomes of seven individuals from basal layers at the site of Xiaohe in the Tarim Basin (Extended Data Table 2). To the best of our knowledge, these individuals represent the earliest human remains excavated to date in the region.

## Genetic diversity of the Bronze Age Xinjiang

We obtained genome-wide data for 18 of 33 attempted individuals by either whole-genome sequencing or DNA enrichment for a panel of about 1.2 million single-nucleotide polymorphisms (1,240k panel SNPs) (Supplementary Data 1A). Overall, endogenous DNA was well preserved with minimal levels of contamination (Extended Data Table 1 and Supplementary Data 1A). To explore the genetic profiles of ancient Xinjiang populations, we first calculated the principal components of present-day Eurasian and Native American populations onto which we projected those of ancient individuals. Ancient Xinjiang individuals form several distinct clusters distributed along principal component 1 (PC1) (Fig. 2), the main principal component that separates eastern and western Eurasian populations. EBA Dzungarian individuals from the sites of Ayituohan and Songshugou near the Altai Mountains (Dzungaria_EBA1) fall close to EBA Afanasievo steppe herders from the Altai–Sayan mountains to the north. Genetic clustering with ADMIXTURE further supports this observation (Extended Data Fig. 3). The contemporaneous individuals from the Nileke site near the Tianshan mountains (Dzungaria_EBA2) are slightly shifted along PC1 towards the later Tarim individuals. In contrast to the EBA Dzungarian individuals, the EMBA individuals from the eastern Tarim sites of Xiaohe and Gumugou (Tarim_EMBA1) form a tight cluster close to pre-Bronze Age central steppe and Siberian individuals who share a high level of ancient North Eurasian (ANE) ancestry (for example, Botai_CA). A contemporaneous individual from the Beifang site (Tarim_EMBA2) in the southern Tarim Basin is slightly displaced from the Tarim_EMBA1 towards EBA individuals from the Baikal region.

## Afanasievo genetic legacy in Dzungaria

Outgroup $f_3$ statistics supports a tight genetic link between the Dzungarian and Tarim groups (Extended Data Fig. 2A). Nevertheless, both of the Dzungarian groups are significantly different from the Tarim groups, showing excess affinity with various western Eurasian populations and sharing fewer alleles with ANE-related groups (Extended Data Fig. 2b,c). To understand this mixed genetic profile, we used qpAdm to explore admixture models of the Dzungarian groups with Tarim_EMBA1 or a terminal Pleistocene individual (AG3) from the Siberian site of Afontova Gora[31], as a source (Supplementary Data 1D). AG3 is a distal representative of the ANE ancestry and shows a high affinity with Tarim_EMBA1. Although the Tarim_EMBA1 individuals lived a millennium later than the Dzungarian groups, they are more genetically distant from the Afanasievo than the Dzungarian groups, suggesting that they have a higher proportion of local autochthonous ancestry. Here we define autochthonous to signify a genetic profile that has been present in a region for millennia, rather than being associated with more recently arrived groups.

We find that Dzungaria_EBA1 and Dzungaria_EBA2 are both best described by three-way admixture models (Fig. 3c, Extended Data Table 3 and Supplementary Data 1D) in which they derive a majority ancestry from Afanasievo (about 70% in Dzungaria_EBA1 and about 50% in Dzungaria_EBA2), with the remaining ancestry best modelled as a mixture of AG3/Tarim_EMBA1 (19–36%) and Baikal_EBA (9–21%). When we use Eneolithic and Bronze Age populations from the IAMC as a source, models fail when Afanasievo is not included as a source, and no contribution is allocated to the IAMC groups when Afanasievo is included (Supplementary Data 1D). Thus, Afanasievo ancestry, without IAMC contributions, is sufficient to explain the western Eurasian component of the Dzungarian individuals. We also find that the Chemurchek, an EBA pastoralist culture that succeeds the Afanasievo in both the Dzungarian Basin and Altai Mountains, derive approximately two-thirds of their ancestry from Dzungaria_EBA1 with the remainder from Tarim_EMBA1 and IAMC/BMAC-related sources (Fig. 3, Extended Data Table 3, Supplementary Data 1F and Supplementary Text 5). This helps to explain both the IAMC/BMAC-related ancestry previously noted in Chemurchek individuals[30] and their reported cultural and genetic affiliations to Afanasievo groups[32]. Taken together, these results indicate that the early dispersal of the Afanasievo herders into Dzungaria was accompanied by a substantial level of genetic mixing with local autochthonous populations, a pattern distinct from that of the initial formation of the Afanasievo culture in southern Siberia.

## Genetic isolation of the Tarim group

The Tarim_EMBA1 and Tarim_EMBA2 groups, although geographically separated by over 600 km of desert, form a homogeneous population that had undergone a substantial population bottleneck, as suggested by their high genetic affinity without close kinship, as well as by the limited diversity in their uniparental haplogroups (Figs. 1 and 2, Extended Data Fig. 4, Extended Data Table 1, Supplementary Data 1B and Supplementary Text 4). Using qpAdm, we modelled the Tarim Basin individuals as a mixture of two ancient autochthonous Asian genetic groups: the ANE, represented by an Upper Palaeolithic individual from the Afontova Gora site in the upper Yenisei River region of Siberia (AG3) (about 72%), and ancient Northeast Asians, represented by Baikal_EBA (about 28%) (Supplementary Data 1E and Fig. 3a). Tarim_EMBA2 from Beifang can also be modelled as a mixture of Tarim_EMBA1 (about 89%) and Baikal_EBA (about 11%). For both Tarim groups, admixture models

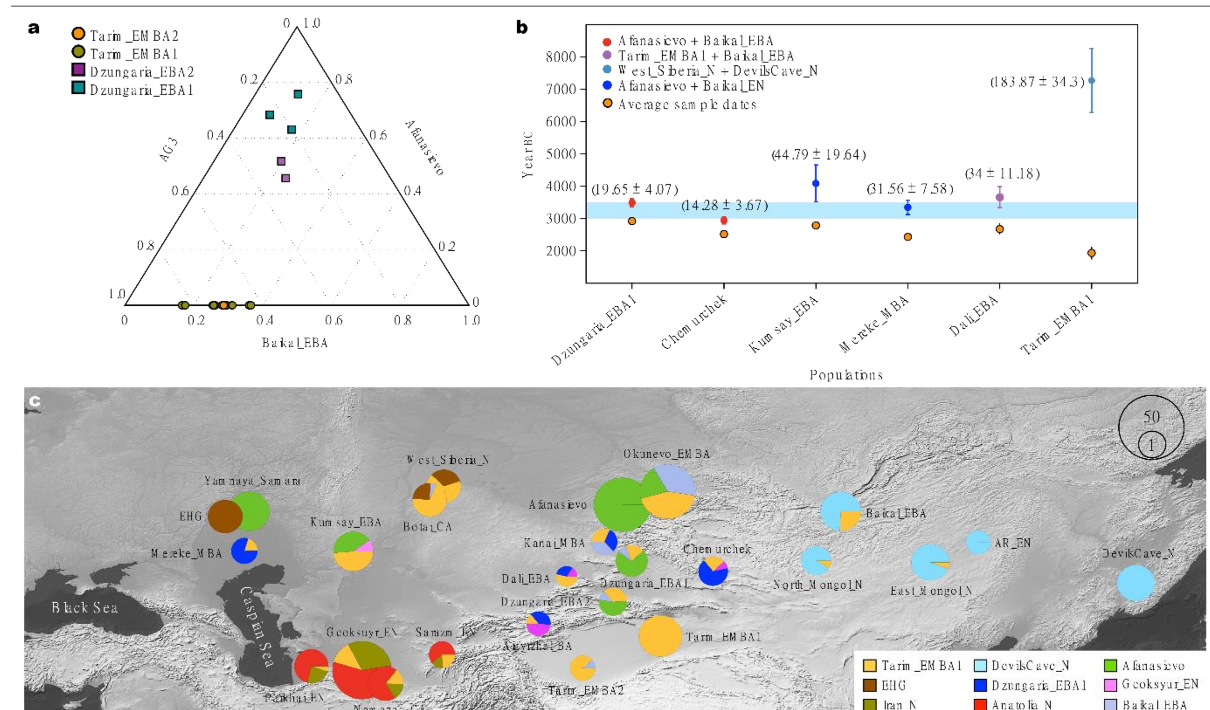


**Fig. 3 | Genetic ancestry and admixture dating of ancient populations from Xinjiang and its vicinity. a**, qpAdm-based estimates of the ancestry proportion of Dzungaria_EBA and Tarim_EMBA from three ancestry sources (AG3, Afanasievo and Baikal_EBA) (Supplementary Data 1D, E). Unlike Dzungaria_EBA individuals, Tarim_EMBA individuals are adequately modelled without EBA Eurasian steppe pastoralist (for example, Afanasievo) ancestry. **b**, Genetic admixture dates for key Bronze Age populations in Inner Asia, including Dzungaria_EBA1 ($n$ = 3), Chemurchek ($n$ = 3), Kumsay_EBA ($n$ = 4), Mereke_MBA ($n$ = 2), Dali_EBA ($n$ = 1) and Tarim_EMBA1 ($n$ = 12). The blue shade represents the radiocarbon dating range of the Yamnaya and Afanasievo individuals. The orange circles and the associated vertical bars represent the averages and standard deviations of median radiocarbon dates, respectively. The circles above each orange circle represent the estimated admixture dates with a generation time of 29 years, and the vertical bars represent the sum of standard errors of the admixture date and the radiocarbon date estimate. **c**, Representative qpAdm-based admixture models of ancient Eurasian groups (Supplementary Data 1D–I). For Dzungaria_EBA1 and Geoksyur_EN, we show their three-way admixture models including Tarim_EMBA1 as a source. For later populations in Xinjiang, IAMC and nearby regions, we used them as sources, and allocated a colour to each of them (blue for Dzungaria_EBA1; magenta for Geoksyur_EN). The basemap in **c** was obtained from the Natural Earth public domain map dataset (https://www.naturalearthdata.com/downloads/10m-raster-data/10m-gray-earth/).

unanimously fail when using the Afanasievo or IAMC/BMAC groups as a western Eurasian source (Supplementary Data 1E), thus rejecting a western Eurasian genetic contribution from nearby groups with herding and/or farming economies. We estimate a deep formation date for the Tarim_EMBA1 genetic profile, consistent with an absence of western Eurasian EBA admixture, placing the origin of this gene pool at 183 generations before the sampled Tarim Basin individuals, or 9,157 ± 986 years ago when assuming an average generation time of 29 years (Fig. 3b). Considering these findings together, the genetic profile of the Tarim Basin individuals indicates that the earliest individuals of the Xiaohe horizon belong to an ancient and isolated autochthonous Asian gene pool. This autochthonous ANE-related gene pool is likely to have formed the genetic substratum of the pre-pastoralist ANE-related populations of Central Asia and southern Siberia (Fig. 3c, Extended Data Fig. 2 and Supplementary Text 5).

## Pastoralism in the Tarim Basin

Although the harsh environment of the Tarim Basin may have served as a strong barrier to gene flow into the region, it was not a barrier to the flow of ideas or technologies, as foreign innovations, such as dairy pastoralism and wheat and millet agriculture, came to form the basis of the Bronze Age Tarim economies. Woollen fabrics, horns and bones of cattle, sheep and goats, livestock manure, and milk and kefir-like dairy products have been recovered from the upper layers of the Xiaohe and Gumugou cemeteries[33–36], as have wheat and millet seeds and bundles of *Ephedra* twigs[34,37,38]. Famously, many of the mummies dating to 1650–1450 bc were even buried with lumps of cheese[35]. However, until now it has not been clear whether this pastoralist lifestyle also characterized the earliest layers at Xiaohe.

To better understand the dietary economy of the earliest archaeological periods, we analysed the dental calculus proteomes of seven individuals at the site of Xiaohe dating to around 2000–1700 bc. All seven individuals were strongly positive for ruminant-milk-specific proteins (Extended Data Table 2), including β-lactoglobulin, α-S1-casein and α-lactalbumin (Extended Data Fig. 5), and peptide recovery was sufficient to provide taxonomically diagnostic matches to cattle (*Bos*), sheep (*Ovis*) and goat (*Capra*) milk (Extended Data Fig. 5, Extended Data Table 2 and Supplementary Data 3). These results confirm that dairy products were consumed by individuals of autochthonous ancestry (Tarim_EMBA1) buried in the lowest levels of the Xiaohe cemetery (Extended Data Table 2). Importantly, however, and in contrast to previous hypotheses[36], none of the Tarim individuals was genetically lactase persistent (Supplementary Data 1J). Rather, the Tarim mummies contribute to a growing body of evidence that prehistoric dairy pastoralism in Inner and East Asia spread independently of lactase persistence genotypes[28,30].

## Discussion

Although human activities in Xinjiang can be traced back to around 40,000 years ago[24,39], the earliest evidence for sustained human habitation in the Tarim Basin dates only to the late third to early second millennium bc. There, at the sites of Xiaohe, Gumugou and Beifang, well-preserved mummified human remains buried within wooden coffins and associated with rich organic grave good assemblages represent the earliest known archaeological cultures of the region. Since their initial discovery in the early twentieth century and subsequent large-scale excavations beginning in the 1990s (ref.[16]), the Tarim mummies have been at the centre of debates with regard to their origins, their relationship to other Bronze Age steppe (Afanasievo), oasis (BMAC) and mountain (IAMC and Chemurchek) groups, and their potential connection to the spread of Indo-European languages into this region[3,4,40].

The palaeogenomic and proteomic data we present here suggest a very different and more complex population history than previously proposed. Although the IAMC may have been a vector for transmitting cultural and economic factors into the Tarim Basin, the known sites from the IAMC do not provide a direct source of ancestry for the Xiaohe populations. Instead, the Tarim mummies belong to an isolated gene pool whose Asian origins can be traced to the early Holocene epoch. This gene pool is likely to have once had a much wider geographic distribution, and it left a substantial genetic footprint in the EMBA populations of the Dzungarian Basin, IAMC and southern Siberia. The Tarim mummies' so-called Western physical features are probably due to their connection to the Pleistocene ANE gene pool, and their extreme genetic isolation differs from the EBA Dzungarian, IAMC and Chemurchek populations, who experienced substantial genetic interactions with the nearby populations mirroring their cultural links, pointing towards a role of extreme environments as a barrier to human migration.

In contrast to their marked genetic isolation, however, the populations of the Xiaohe horizon were culturally cosmopolitan, incorporating diverse economic elements and technologies with far-flung origins. They made cheese from ruminant milk using a kefir-like fermentation[37], perhaps learned from descendants of the Afanasievo, and they cultivated wheat, barley and millet[37,41], crops that were originally domesticated in the Near East and northern China and which were introduced into Xinjiang no earlier than 3500 bc (refs.[8,42]), probably via their IAMC neighbours[24]. They buried their dead with *Ephedra* twigs in a style reminiscent of the BMAC oasis cultures of Central Asia, and they also developed distinctive cultural elements not found among other cultures in Xinjiang or elsewhere, such as boat-shaped wooden coffins covered with cattle hides and marked by timber poles or oars, as well as an apparent preference for woven baskets over pottery[43,44]. Considering these findings together, it appears that the tight knit population that founded the Xiaohe horizon were well aware of different technologies and cultures outside the Tarim Basin and that they developed their unique culture in response to the extreme challenges of the Taklamakan Desert and its lush and fertile riverine oases[4].

This study illuminates in detail the origins of the Bronze Age human populations in the Dzungarian and Tarim basins of Xinjiang. Notably, our results support no hypothesis involving substantial human migration from steppe or mountain agropastoralists for the origin of the Bronze Age Tarim mummies, but rather we find that the Tarim mummies represent a culturally cosmopolitan but genetically isolated autochthonous population. This finding is consistent with earlier arguments that the IAMC served as a geographic corridor and vector for regional cultural interaction that connected disparate populations from the fourth to the second millennium bc (refs.[24,25]). While the arrival and admixture of Afanasievo populations in the Dzungarian Basin of northern Xinjiang around 3000 bc may have plausibly introduced Indo-European languages to the region, the material culture and genetic profile of the Tarim mummies from around 2100 bc onwards call into question simplistic assumptions about the link between genetics, culture and language and leave unanswered the question of whether the Bronze Age Tarim populations spoke a form of proto-Tocharian. Future archaeological and palaeogenomic research on subsequent Tarim Basin populations—and most importantly, studies of the sites and periods where first millennium ad Tocharian texts have been recovered—are necessary to understand the later population history of the Tarim Basin. Finally, the palaeogenomic characterization of the Tarim mummies has unexpectedly revealed one of the few known Holocene-era genetic descendant populations of the once widespread Pleistocene ANE ancestry profile. The Tarim mummy genomes thus provide a critical reference point for genetically modelling Holocene-era populations and reconstructing the population history of Asia.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04052-7.


1. Peyrot, M. in *Aspects of Globalisation: Mobility, Exchange and the Development of Multi-Cultural States* 12–17 (2017).
2. Damgaard, P. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
3. Hemphill, B. E. & Mallory, J. P. Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. *Am. J. Phys. Anthropol.* **124**, 199–222 (2004).
4. Betts, A., Jia, P. & Abuduresule, I. A new hypothesis for early Bronze Age cultural diversity in Xinjiang, China. *Archaeol. Res. Asia* **17**, 204–213 (2019).
5. Li, C. et al. Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol.* **8**, 15 (2010).
6. Li, C. et al. Analysis of ancient human mitochondrial DNA from the Xiaohe cemetery: insights into prehistoric population movements in the Tarim Basin, China. *BMC Genet.* **16**, 78 (2015).
7. Ning, C. et al. Ancient genomes reveal Yamnaya-related ancestry and a potential source of Indo-European speakers in Iron Age Tianshan. *Curr. Biol.* **29**, 2526–2532 (2019).
8. Zhou, X. et al. 5,200-year-old cereal grains from the eastern Altai Mountains redate the trans-Eurasian crop exchange. *Nat. Plants* **6**, 78–87 (2020).
9. Wang, T. et al. Tianshanbeilu and the isotopic millet road: reviewing the late Neolithic/Bronze Age radiation of human millet consumption from north China to Europe. *Natl Sci. Rev.* **6**, 1024–1039 (2019).
10. Zhang, Y. et al. Holocene environmental changes around Xiaohe Cemetery and its effects on human occupation, Xinjiang, China. *J. Geogr. Sci.* **27**, 752–768 (2017).
11. Hong, Z., Jian-Wei, W., Qiu-Hong, Z. & Yun-Jiang, Y. A preliminary study of oasis evolution in the Tarim Basin, Xinjiang, China. *J. Arid Environ.* **55**, 545–553 (2003).
12. Jia, P. & Betts, A. A re-analysis of the Qiemu'erqieke (Shamirshak) cemeteries, Xinjiang, China. *J. Indo-Eur. Stud.* **38**, 275–317 (2010).
13. Peyrot, M. The deviant typological profile of the Tocharian branch of Indo-European may be due to Uralic substrate influence. *Indo-Eur. Linguist.* **7**, 72–121 (2019).
14. Bouckaert, R. et al. Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957–960 (2012).
15. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
16. Mallory, J. P. & Mair, V. H. *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West* (Thames & Hudson, 2000).
17. Barber, E. W. *Mummies of Urumchi* (W. W. Norton & Co., 1999).
18. Mair, V. H. Prehistoric Caucasoid corpses of the Tarim Basin. *J. Indo-Euro. Stud.* **23**, 281–307 (1995).
19. Mair, V. H. in *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia* Vol. 2 835–855 (Institute for the Study of Man and the University of Pennsylvania Museum, 1998).
20. Mallory, J. P. *The Problem of Tocharian Origins: an Archaeological Perspective* (Univ. Pennsylvania Press, 2015).
21. Chen, K. & Hiebert, F. T. The late prehistory of Xinjiang in relation to its neighbors. *J. World Prehist.* **9**, 243–300 (1995).
22. Han, K. Craniometric study on the ancient individuals from the Gumugou site, Xinjiang (in Chinese). *Kaogu Xuebao* 361–384 (1986).
23. Kuzmina, E. E. in *Archeology, Migration and Nomadism, Linguistics* Vol. 1 63–93 (Univ. Pennsylvania Museum Publications, 1998).
24. Li, Y. Agriculture and palaeoeconomy in prehistoric Xinjiang, China (3000–200 BC). *Veg. Hist. Archaeobot.* **30**, 287–303 (2021).
25. Frachetti, M. D. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr. Anthropol.* **53**, 2–38 (2012).
26. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
27. Feng, Q. et al. Genetic history of Xinjiang's Uyghurs suggests Bronze Age multiple-way contacts in Eurasia. *Mol. Biol. Evol.* **34**, 2572–2582 (2017).

28. Jeong, C. et al. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc. Natl Acad. Sci. USA* **115**, E11248–E11255.
29. Yu, H. et al. Paleolithic to Bronze Age Siberians reveal connections with first Americans and across Eurasia. *Cell* **181**, 1232–1245 (2020).
30. Jeong, C. et al. A dynamic 6,000-year genetic history of Eurasia's Eastern Steppe. *Cell* **183**, 890–904 (2020).
31. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
32. Wang, C.-C. et al. Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021).
33. Li, J.-F. et al. Buried in sands: environmental analysis at the archaeological site of Xiaohe Cemetery, Xinjiang, China. *PLoS ONE* **8**, e68957 (2013).
34. Qiu, Z. et al. Paleo-environment and paleo-diet inferred from Early Bronze Age cow dung at Xiaohe Cemetery, Xinjiang, NW China. *Quat. Int.* **349**, 167–177 (2014).
35. Yang, Y. et al. Proteomics evidence for kefir dairy in Early Bronze Age China. *J. Archaeol. Sci.* **45**, 178–186 (2014).
36. Xie, M. et al. Identification of a dairy product in the grass woven basket from Gumugou Cemetery (3800 BP, northwestern China). *Quat. Int.* **426**, 158–165 (2016).
37. Yang, R. et al. Investigation of cereal remains at the Xiaohe Cemetery in Xinjiang, China. *J. Archaeol. Sci.* **49**, 42–47 (2014).
38. Zhang, G. et al. Ancient plant use and palaeoenvironmental analysis at the Gumugou Cemetery, Xinjiang, China: implication from desiccated plant remains. *Archaeol. Anthropol. Sci.* **9**, 145–152 (2017).
39. Yu, J. & He, J. Significant discoveries from the excavation of Jimunai Tongtiandong site (in Chinese). *Wenwubao* **8** (2017).
40. Hollard, C. et al. New genetic evidence of affinities and discontinuities between Bronze Age Siberian populations. *Am. J. Phys. Anthropol.* **167**, 97–107 (2018).
41. Li, C. et al. Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. *J. Archaeol. Sci.* **38**, 115–119 (2011).
42. Stevens, C. J. & Fuller, D. Q. The spread of agriculture in eastern Asia: archaeological bases for hypothetical farmer/language dispersals. *Lang. Dyn. Change* **7**, 152–186 (2017).
43. Abuduresule, I. Archaeological report of Xiaohe cemetery of 2003 (in Chinese). *Wenwu* 4–42 (2007).
44. Abuduresule, Y., Li, W. & Hu, X. in *The Cultures of Ancient Xinjiang, Western China: Crossroads of the Silk Roads* 19–51 (Archaeopress, 2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

## Methods

### Sample provenance

The archaeological human remains studied in this manuscript were excavated by the Xinjiang Institute of Cultural Relics and Archaeology from 1979 to 2017. Scientific investigation of these remains was approved by the Xinjiang Cultural Relics and Archaeology Institute, which holds the custodianship of the studied remains, based on the written agreements.

### Radiocarbon dating

Of the 18 individuals reported in this study, 10 were directly dated using accelerator mass spectrometry (AMS) at Beta Analytic, Miami, USA, and/or at Lanzhou University, China. To confirm the reliability of our AMS dating results, 4 out of the 10 individuals were AMS-dated at both Beta Analytic and Lanzhou University. Consistent dates were obtained in all cases (Supplementary Data 1C). The calibration of the dated samples was performed on the basis of the IntCal20 database[45] and using the OxCal v.4.4 program[46]. All of the samples were dated to time periods consistent with those estimated from archaeological stratigraphic layers and excavated grave goods.

### DNA laboratory procedures

Ancient DNA work was conducted in dedicated clean room laboratory facilities at the ancient DNA laboratories of Jilin University in Changchun and the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing (Extended Data Table 1 and Supplementary Data 1A). For the 33 individuals initially screened in this study, approximately 50 mg of dentine or bone powder was obtained per individual from either teeth or bones. DNA was extracted following established protocols[47] with slight modifications (https://doi.org/10.17504/protocols.io.baksicwe). A subset of DNA extracts ($n$ = 16) was subjected to a partial uracil-specific excision reagent repair following the methods described in ref.[48] (Extended Data Table 1 and Supplementary Data 1A). All 33 DNA extracts were built into double-stranded dual-index Illumina libraries. Libraries that were prepared in Jilin ($n$ = 26) were directly shotgun sequenced on an Illumina HiSeq X10 or HiSeq 4000 instrument using 2 × 150-base-pair (bp) chemistry, and those with endogenous human DNA higher than 10% ($n$ = 12) were sent for deeper sequencing. One of the 12 individuals (XHBM1) was later excluded from this study owing to high modern human DNA contamination (Supplementary Data 1A). For libraries prepared at the Institute of Vertebrate Paleontology and Paleoanthropology, samples with 0.1% or more human DNA from the initial screening ($n$ = 7) were further enriched for approximately 1.2 million nuclear SNPs and then deeper sequenced on an Illumina HiSeq 4000 instrument using 2 × 150-bp chemistry. Together, a total of 18 individuals yielded sufficient high-quality ancient genomic data for downstream analyses (Extended Data Table 1).

### DNA sequence data processing

Raw read data were processed with EAGER v.1.92.55 (ref.[49]), a pipeline specially designed for processing ancient DNA sequence data. Specifically, raw reads were trimmed for Illumina adaptor sequences, and overlapping pairs were collapsed into single reads using AdapterRemoval 2.2.0 (ref.[50]). Merged reads were mapped to the human reference genome (hs37d5; GRCh37 with decoy sequences) using the aln/samse programs in BWA v.0.7.12 (ref.[51]). PCR duplicates were removed using DeDup v.0.12.2 (ref.[49]). To minimize the effect of postmortem DNA damage on genotyping, we trimmed BAM files generated from samples treated ($n$ = 11) or not ($n$ = 7) with uracil DNA glycosylase (UDG) by soft-masking up to 10 bp on both ends of each read using the trimbam function on bamUtils v.1.0.13 (ref.[52]) on the basis of the DNA misincorporation pattern per library tabulated using mapDamage v.2.0.9 (ref.[53]). For each SNP in the 1,240k panel, a single base from a high-quality read (base and mapping quality score 30 or higher) was randomly sampled to represent a pseudo-diploid genotype using the pileupCaller v.1.4.0.5 downloaded from https://github.com/stschiff/sequenceTools under the random haploid calling mode (-randomHaploid). For the transition SNPs (C/T and G/A), trimmed BAM files were used. For the transversion SNPs, BAM files without trimming were used.

### Ancient DNA authentication

We assessed the authenticity of our ancient DNA data as follows. First, we computed the proportion of C-to-T deamination errors at both the 5′ and 3′ ends of the sequencing reads, and found that all samples exhibited postmortem damage patterns characteristic of ancient DNA (Supplementary Data 1A). We then estimated mitochondrial DNA contamination for all individuals using the Schmutzi v.1.5.1 program[54]. To do this, we mapped adapter-trimmed reads to a 500-bp-extended revised Cambridge Reference Sequence (rCRS) of the human mitochondrial genome (NC_012920.1) to preserve reads passing through the origin, and then wrapped up the alignment to the regular rCRS with the circularmapper v.1.1 (ref.[49]). We successively ran the contDeam and schmutzi modules in the schmutzi program against the worldwide allele frequency database of 197 individuals to estimate the mitochondrial DNA contamination rate. Last, we estimated the nuclear contamination rate on men using ANGSD v.0.910 (ref.[55]), on the basis of the principle that men have only a single copy of the X chromosome, and thus contamination will introduce extra mismatches among reads in SNP sites but not in the flanking monomorphic sites.

### DNA reference datasets

We compared the genome sequences of our ancient individuals to two sets of worldwide genotype panels, one based on the Affymetrix Axiom Genome-wide Human Origins 1 array (HumanOrigins; 593,124 autosomal SNPs)[56–58] and the other on the 1,240k dataset (1,233,013 autosomal SNPs including all of the HumanOrigins SNPs)[59]. We augmented both datasets by adding the Simons Genome Diversity Panel[60] and published ancient genomes (Supplementary Data 2A).

### Genetic relatedness analysis

We used pairwise mismatch rate (pmr)[61] and lcMLkin v0.5.0 (ref.[62]), to determine the genetic relatedness between ancient individuals. We calculated pmr for all pairs of ancient individuals in this study using the autosomal SNPs in the 1,240k panel and kept individual pairs with at least 8,000 SNPs covered by both to remove noisy estimates from low-coverage samples. We used lcMLkin to validate our observation in pmr analysis and to distinguish between parent–offspring and full sibling pairs.

### Uniparental haplogroup assignment

We aligned the adapter-trimmed reads to the rCRS NC_012920.1, and then generated the mitochondrial consensus sequence of each ancient individual using Geneious software v.11.1.3 (ref.[63]; https://www.geneious.com/). We assigned each consensus sequence into a specific haplogroup using HaploGrep2 (ref.[64]). For the Y chromosome, we used lineage-informative SNPs from the International Society of Genetic Genealogy 2016 tree (https://isogg.org/tree/2016/index16.html). For these SNPs, we called each individual's genotype using bcftools v.1.7 (ref.[51]) mpileup and call modules, after removing reads with mapping quality score < 30 (-q 30) and bases with quality score < 30 (-Q 30). We subsequently removed all heterozygous genotype calls. Then we assigned each individual to a specific Y haplogroup by manually comparing the genotype calls with the International Society of Genetic Genealogy SNPs. Before variant calling, we filtered alignment data using the pysam library v.0.15.2 (https://pysam.readthedocs.io/en/latest/) to reduce false positive variants due to postmortem damage and modern human contamination. We kept an observed base only if it was from a read shorter than 100 bp and the base was more than 10 bp away from the read ends. For transition SNPs, we further removed aligned bases if they were from a read with no postmortem damage pattern (that is, no C-to-T or G-to-A substitution). We determined each individual's Y

haplogroup primarily on the basis of the transversion SNPs and additionally considered transitions if transversions were insufficient.

### Population genetic analysis

We performed principal component analysis as implemented in smartpca v.16000 (ref. [65]) using a set of 2,077 present-day Eurasian individuals from the Human Origins dataset (Supplementary Data 2B) with the options 'lsqproject: YES' and 'shrinkmode: YES'. The unsupervised admixture analysis was performed with ADMIXTURE v.1.3.0 (ref. [66]). For ADMIXTURE, we removed genetic markers with minor allele frequency lower than 1% and pruned for linkage disequilibrium using the -indep-pairwise 200 25 0.2 option in PLINK v.1.90 (ref. [67]). We used outgroup $f_3$ statistics[68] to obtain a measurement of genetic relationship of the target population to a set of the Eurasian populations since their divergence from an African outgroup. We calculated $f_4$ statistics with the 'f4mode: YES' function in the ADMIXTOOLS package[58]. $f_3$ and $f_4$ statistics were calculated using qp3Pop v.435 and qpDstat v.755 in the ADMIXTOOLS package.

### Runs of homozygosity

We characterized whether the Bronze Age Xinjiang individuals descended from genetically related parents by estimating the runs of homozygosity (ROH). ROH refers to segments of the genome where the two chromosomes in an individual are identical to each other owing to recent common ancestry. Therefore, the presence of long ROH segments strongly suggests that an individual's parents are related. We applied the hapROH method[69] using the Python library hapROH v.0.3a4 with default parameters. The method was developed to identify ROH from low-coverage genotype data typical of ancient DNA and is still robust enough to identify ROH for individuals with a coverage down to 0.5× (ref. [69]). We reported the total sum of ROH longer than 4, 8, 12 and 20 cM, and visualized the results using DataGraph v.4.5.1.

### Genetic admixture modelling with qpAdm

We modelled our ancient Xinjiang populations using the qpWave/qpAdm programs (qpWave v.410 (ref. [70]) and qpAdm v.810 (ref. [57])). We used the following eight populations in the 1,240k dataset as the base set of outgroups (base) unless explicitly stated otherwise: Mbuti ($n$ = 5), Natufian ($n$ = 6), Onge ($n$ = 2), Iran_N ($n$ = 5), Villabruna ($n$ = 1), Mixe ($n$ = 3), Ami ($n$ = 2), Anatolia_N ($n$ = 23). This set includes an African outgroup (Mbuti), early Holocene Levantine hunter-gatherers (Natufian), Andamanese islanders (Onge), early Neolithic Iranians from the Tepe Ganj Darch site (Iran_N), late Pleistocene Western European hunter-gatherers (Villabruna), Central Native Americans (Mixe), an indigenous group native to Taiwan (Ami) and Neolithic farmers from Anatolia (Anatolia_N). To compare competing models, we also took a 'rotating' approach, where we reciprocally added a source from a model to outgroups for a competing model. We specified which outgroups are used for all qpAdm models.

### Admixture dating with DATES

We used DATES v.753 (ref. [26]) for the dating of admixture events of the ancient populations with the pseudo-haploid genotype data under the simplified assumption that gene flow occurred as a single event, and assuming a generation time of 29 years (ref. [58]). The DATES software measures the decay of ancestry covariance to infer the admixture time and estimates jackknife standard errors. In the parameter file for running DATES, we used the options binsize: 0.001, maxdis: 0.5, runmode: 1, qbin: 10 and lovalfit: 0.45 in every run on the pseudo-haploid genotype data. For each target population, we chose a pair of reference populations that we identified as good sources in the qpAdm analysis. In cases in which the qpAdm source had limited sample size or SNP coverage, we chose an alternative that had a similar genetic profile to the qpAdm source but with better data quality to enhance the statistical power of the DATES analysis (Supplementary Data 1D–G). For Dzungaria_EBA1 and Chemurchek, we used the Afanasievo ($n$ = 20) and Baikal_EBA ($n$ = 9) as the references. For Kumsay_EBA and Mereke_MBA, we used the Afanasievo ($n$ = 20) and Baikal_EN ($n$ = 15). For Dali_EBA, we used Tarim_EMBA1 ($n$ = 12) and Baikal_EBA ($n$ = 9). For Tarim_EMBA1, we used West_Siberia_N ($n$ = 3) and DevilsCave_N ($n$ = 4).

### Protein extraction, digestion and liquid chromatography with tandem mass spectrometry

Total protein extractions were performed on dental calculus obtained from seven Xiaohe individuals excavated from layers 4 and 5 (Extended Data Table 2). Only individuals with calculus deposits >5 mg were analysed, and 5–10 mg of dental calculus was processed for each sample. Samples were extracted and digested using a filter-aided sample preparation, following decalcification in 0.5 M EDTA (ref. [71]). Extracted peptides were analysed by liquid chromatography with tandem mass spectrometry (MS/MS) using a Q-Exactive mass spectrometer (Thermo Scientific) coupled to an ACQUITY UPLC M-Class system (Waters AG) according to previously described protocols[28]. Potential contamination and sample carryover were monitored through the use of extraction blanks as well as injection blanks between each sample.

### Protein database searching

Tandem mass spectra were converted to Mascot generic files by MSConvert version 3.0.11781 using the 100 most intense MS/MS peaks. All MS/MS samples were analysed using Mascot (Matrix Science; v.2.6.0). Mascot was set up to search the SwissProt Release 2019_08 database (560,823 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.050 Da and a parent ion tolerance of 10.0 ppm. Carbamidomethylation of cysteine was specified in Mascot as a fixed modification. Deamidation of asparagine and glutamine and oxidation of methionine and proline were specified in Mascot as variable modifications. A subset of samples were analysed in duplicate (Supplementary Data 3), and the results were combined using multidimensional protein identification technology (MudPIT) before analysis.

### Criteria for protein identification

MS/MS-based protein and peptide identifications were validated using Scaffold (version Scaffold_4.9.0, Proteome Software). Peptide identifications were accepted if they could be established at greater than 86.0% probability to achieve a false discovery rate (FDR) less than 1.0% by the Peptide Prophet algorithm[71] with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at an FDR of less than 5.0% and contained at least two unique peptides. Final protein and peptide FDRs were 1.8% and 0.99%, respectively. Protein probabilities were assigned by the Protein Prophet algorithm[72]. After establishing the presence of the milk proteins β-lactoglobulin and α-S1-casein using these criteria, we expanded our analysis to accept further milk proteins identified on the basis of single peptides for high-scoring PSMs (>60), which resulted in the additional identification of α-lactalbumin. Proteins that contained similar peptides that could not be differentiated on the basis of MS/MS analysis alone were grouped to satisfy the principles of parsimony. All samples yielded proteomes typical of a dental calculus oral microbiome, and damage-associated modifications (N and Q deamidation) characteristic of ancient proteins were observed (Supplementary Data 3).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The DNA sequences reported in this paper have been deposited in the European Nucleotide Archive under the accession number PRJEB46875.

Haploid genotype data of ancient individuals in this study on the 1,240k panel are available in the EIGENSTRAT format at https://edmond.mpdl.mpg.de/imeji/collection/OMm2fpu0jR3jSqnY. The protein spectra have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository under the accession number PDX027706. The publicly available database SwissProt release 2019_08 is accessible through the UniProtKnowledgeBase (https://www.uniprot.org). The basemaps used in Figs. 1, 3 are in the public domain and accessible through the NaturalEarth website (https://www.naturalearthdata.com/downloads/10m-raster-data/).

## Code availability

All of the analyses performed in this study are based on publicly available software programs. Specific version information and non-default arguments are described in the Methods.


45. Reimer, P. J. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
46. Ramsey, C. B. Methods for summarizing radiocarbon datasets. *Radiocarbon* **59**, 1809–1833 (2017).
47. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
48. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B* **370**, (2015).
49. Peltzer, A. et al. EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
50. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, (2016).
51. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
52. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).
53. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
54. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
55. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
56. Jeong, C. et al. The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* **3**, 966–976 (2019).
57. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
58. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
59. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
60. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
61. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
62. Lipatov, M., Sanjeev, K., Patro, R. & Veeramah, K. R. Maximum likelihood estimation of biological relatedness from low coverage sequencing data. Preprint at https://doi.org/10.1101/023374 (2015).
63. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
64. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
65. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
66. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
67. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).
68. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
69. Ringbauer, H., Novembre, J. & Steinrücken, M. Detecting runs of homozygosity from low-coverage ancient DNA. Preprint at https://doi.org/10.1101/2020.05.31.126912 (2020).
70. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
71. Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
72. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).



**Acknowledgements** We thank Xinjiang Institute of Cultural Relics and Archaeology and Renmin University of China for providing such valuable samples for study; Lanzhou University for providing the AMS dating results; K. Wang, H. Yu and G. A. Gnecchi-Ruscone for helpful comments on the genetic landscapes of the Eurasian steppe; T. Hermes and R. Flad for helpful comments on the broader archaeological context of the region. This work was supported by the National Key R&D Program of China (grant numbers 2016YFE0203700 and 2018YFA0606402), the National Natural Science Foundation of China (grant number 42072018, 41925009), the Fundamental Research Funds for the Central Universities, the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement numbers 804884-DAIRYCULTURES and 646612-Eurasia3angle), the Humanities and Social Sciences Key Research Base of the Ministry of Education (16JJD780005), the National Research Foundation of Korea grant funded by the Korean Government (MSIT; 2020R1C1C1003879) and the Max Planck Society.


**Author contributions** Y.C., C.J., C.W., C.N. and J.K. conceived and supervised the study. F.Z., A.S., L.F., P.C., R.Y., F.L. and Q.D. performed the laboratory work. Q.F., D.W., W.L., X.H., Q.R., I.A., C.L., S.G., Y.X., S.Wu, S.Wen, H.Zhu, H.Zhou and A.N. provided archaeological materials and associated information. R.B. and M.R. provided the linguistic background and G.D. and Z.T. assisted with the AMS dating. C.N., F.Z., A.S., C.W., C.J., Q.F., P.M., X.F., W.W. and V.K. analysed data. C.N., C.W., C.J., Y.C., F.Z. and A.S. wrote the manuscript with input from all coauthors.


**Funding** Open access funding provided by Max Planck Society.


**Competing interests** The authors declare no competing interests.

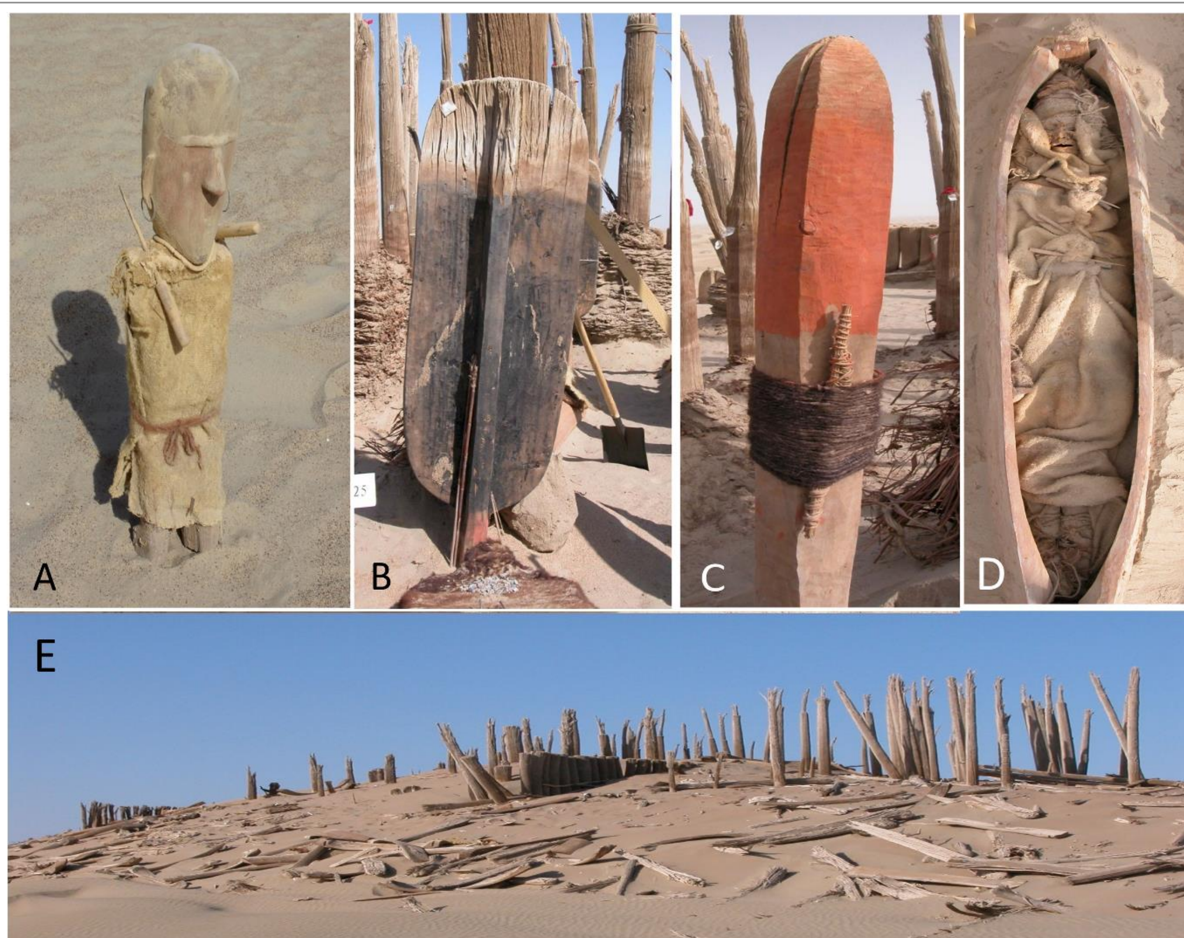
**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04052-7.

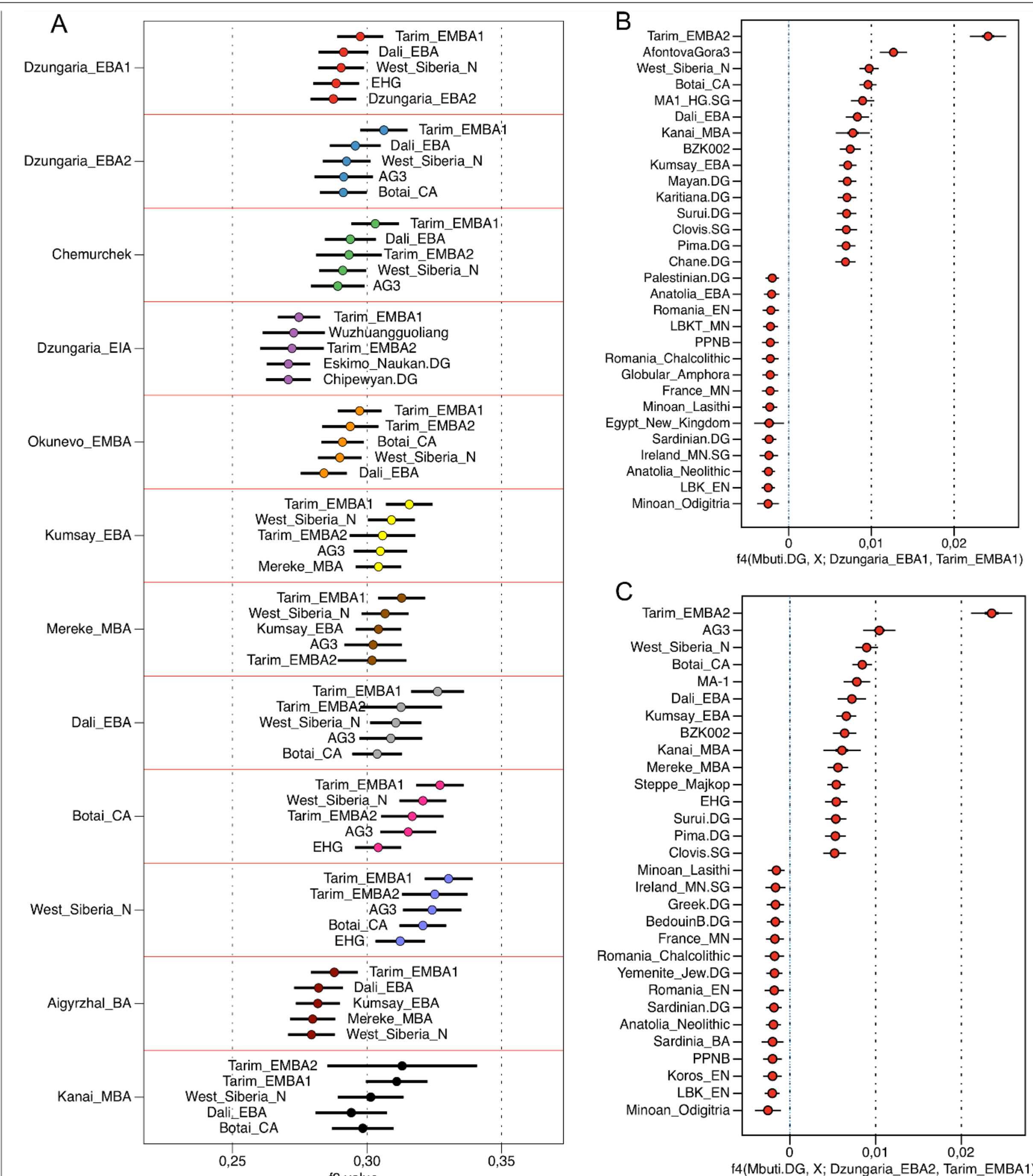

**Correspondence and requests for materials** should be addressed to Chao Ning, Johannes Krause, Christina Warinner, Choongwon Jeong or Yinqiu Cui.


**Peer review information** *Nature* thanks Paula Dupuy, Michael Frachetti and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

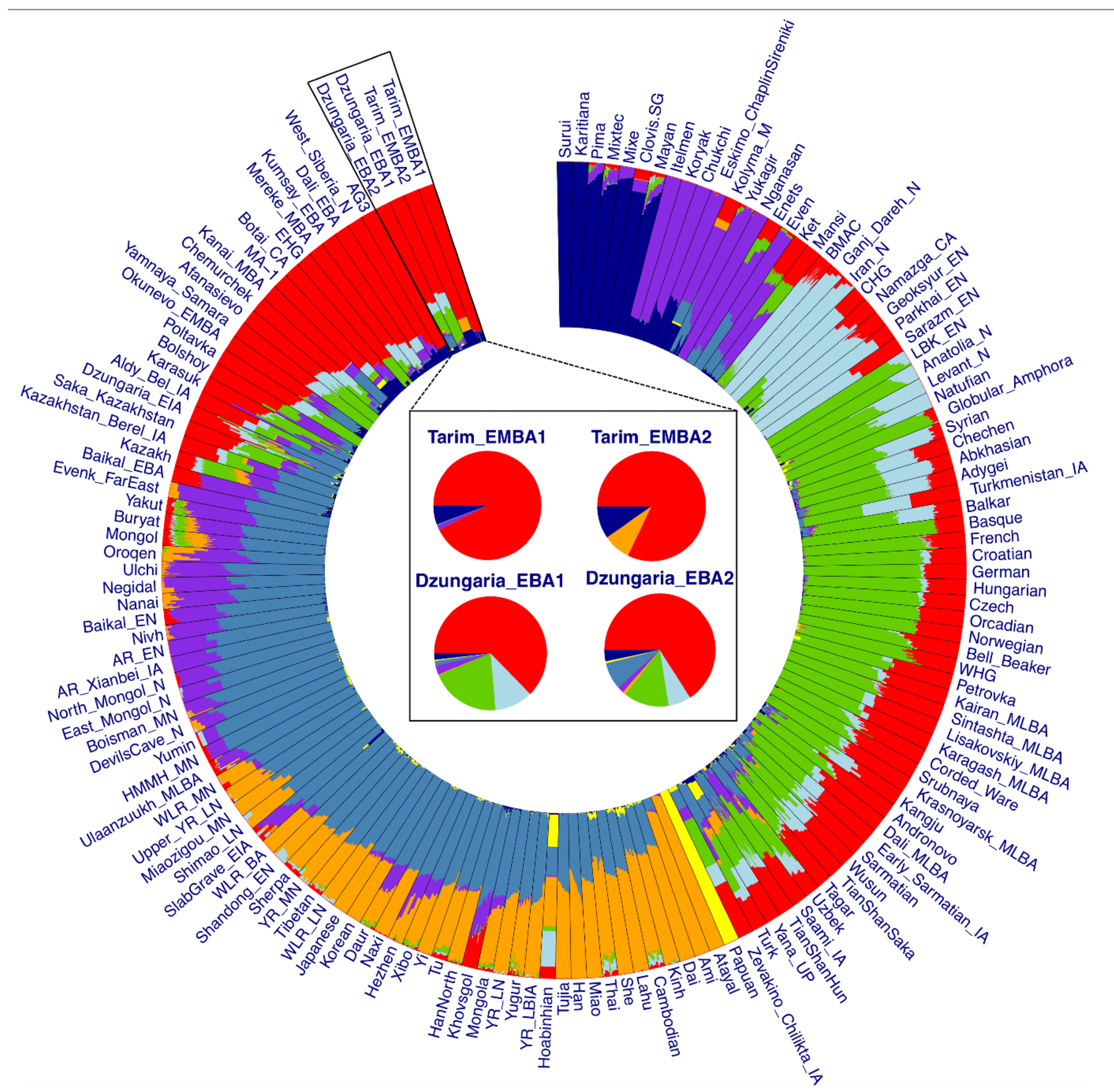

**Reprints and permissions information** is available at http://www.nature.com/reprints.



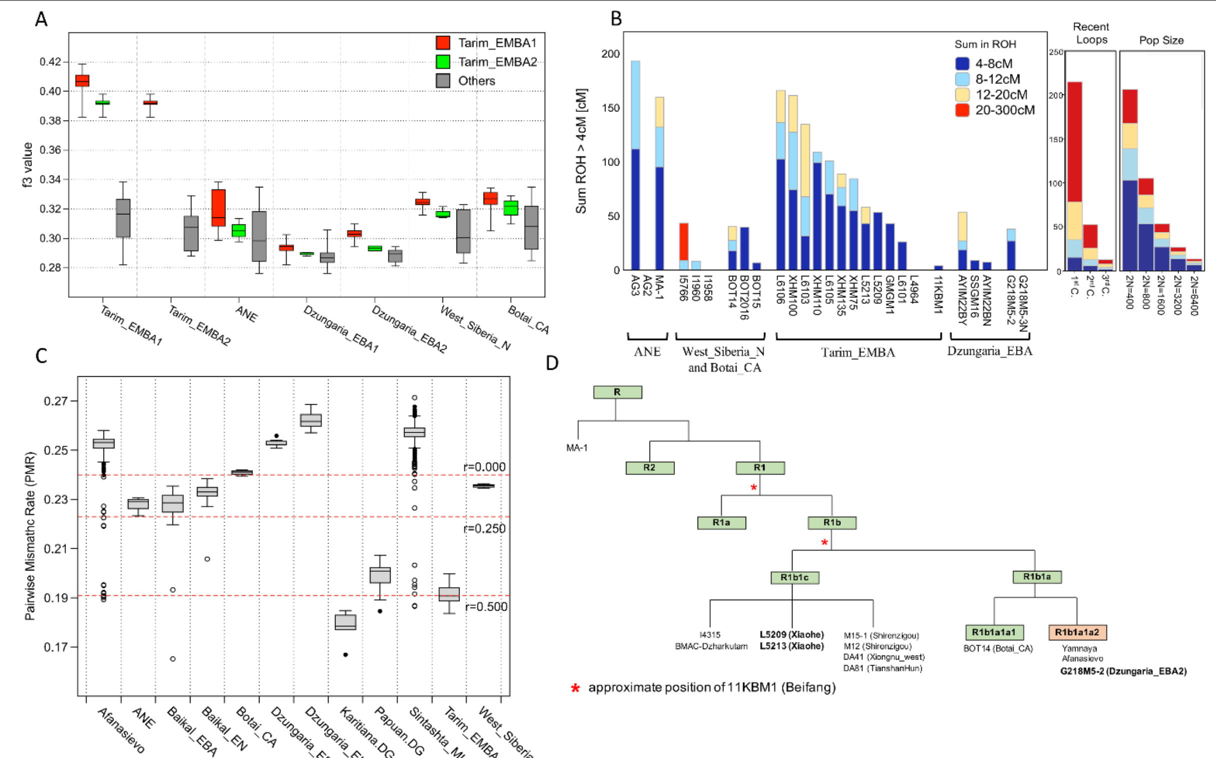

**Extended Data Fig. 1 | Burial goods excavated from the Xiaohe cemetery.** **A**, a wooden sculpture excavated from the upper layer of a double-layer mud coffin of XHM75. **B**, an oar-plank placed in front of a male burial. **C**, a wooden pole placed in front of a female burial. **D**, Burial XHM66 from layer 4 of the Xiaohe cemetery illustrating typical features of early burials, including boat-shaped coffins and mummified remains dressed in woollen garments. This burial style is common at Bronze Age cemeteries throughout the Tarim Basin, including Beifang and Gumugou. **E**, Side view of the Xiaohe cemetery showing wooden grave markers and fencing.



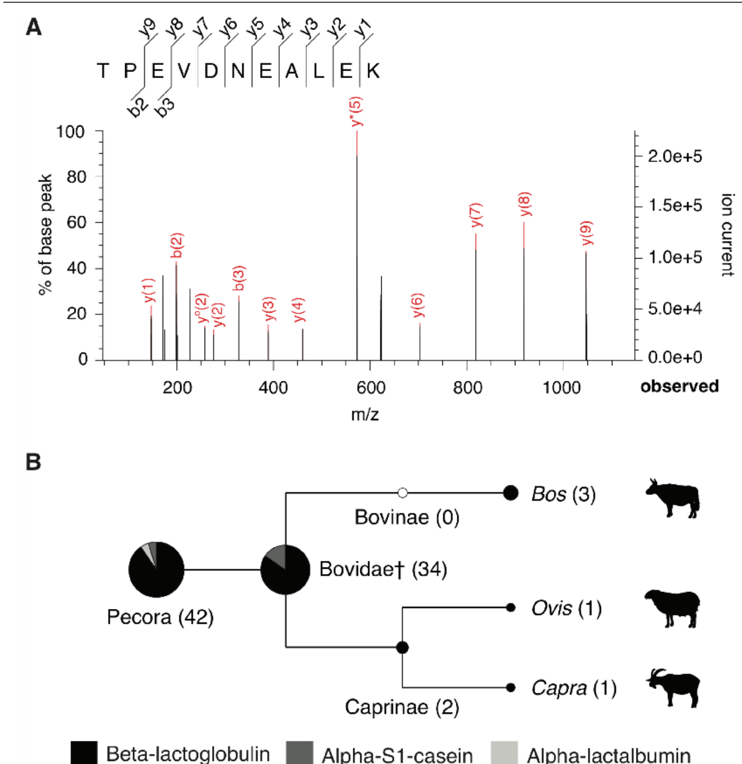

**Extended Data Fig. 2 | F-statistics for the ancient Xinjiang and the Eurasian steppe populations. A**, we show top 5 outgroup *f*3-statistics of the form *f*3(Target, X; Mbuti) for the 361 world-wide populations as contrast populations X, and 8 populations from this study and the Eurasian Steppe as target: Dzungaria_EBA1, Dzungaria_EBA2, Chemurchek, Dzungaria_EIA, Okunevo_EMBA, Kazakhstan_EMBA, Botai_CA, West_Siberia_N, horizontal bars represent ± 1 standard error measure (s.e.m.) calculated by 5cM block jackknifing. **B**, *f*4-statistics of the form *f*4(Mbuti, X; Dzungaria_EBA1, Tarim_EMBA1), horizontal bars represent ± 3 (thin) and ± 1 (thick) s.e.m. calculated by 5cM block jackknifing, and **C**, *f*4-statistics of the form *f*4(Mbuti, X; Dzungaria_EBA2, Tarim_EMBA1), where X is 361 world-wide populations. We show the top and the bottom 15 *f*4 statistics. Horizonal bars represent the point estimate ± 3 (thin) and ± 1 (thick) s.e.m., respectively, as estimated using 5cM block jackknifing. *F4* statistics deviating three s.e.m. or more from zero are marked in red.




**Extended Data Fig. 3 | Unsupervised ADMIXTURE plot for the Bronze Age Xinjiang individuals.** We plot ancestry component estimates for K = 8 using 'AncestryPainter' (https://www.picb.ac.cn/PGG/resource.php). Dzungaria_EBA individuals show an ancestry pattern close to Afanasievo and Yamnaya, while Tarim_EMBA individuals show a pattern similar to AG3, West_Siberia_N and Botai_CA from the Eurasia steppe.




**Extended Data Fig. 4 | Reduced genetic diversity of the Tarim_EMBA individuals. A**, a comparison of individual outgroup *f*3-statistics for the ancient Xinjiang populations and their neighboring populations from Inner Asia, including Tarim_EMBA1 (n = 12), Tarim_EMBA2 (n = 1), ANE (n = 3), Dzungaria_EBA1 (n = 3), Dzungaria_EBA2 (n = 2), West_Siberia_N (n = 3) and Botai_CA (n = 3), which Tarim Basin individuals show the highest affinity to each other. In each boxplot, the box marks the 25th and 75th quartiles of the distribution, respectively, and the horizontal line within the box marks the median. The whisker delineates the maximum and the minimum. **B**, the cumulative distribution of ROH tracts shows that Tarim_EMBA individuals did not descend from close related parents. **C**, pairwise mismatch rate (pmr) between individuals in the ancient populations of Xinjiang and its neighboring regions, including all pairs of individuals within the Afanasievo (n = 27), ANE (n = 3), Baikal_EBA (n = 9), Baikal_EN (n = 15), Botai_CA (n = 3), Dzungaria_EBA (n = 5), Dzungaria_EIA (n = 10), Sintashta_MLBA (n = 51), Tarim_EMBA (n = 13), West_Siberia_N (n = 3), as well as present-day isolated populations such as Papuan and Karitiana. Tarim_EMBA individuals uniformly show a much reduced pmr value that is equivalent to the first-degree relatives in Afanasievo or Sintashta_MLBA. The red dotted lines mark the expected pmr value for the given coefficient of relationship ($r$), ranging from 0 (unrelated) and 1/4 (second degree relatives) to 1/2 (first degree relatives), based on the mean value of pmr among these populations, respectively. In each box plot, the box represents the interquartile range (the 25th and 75th quartiles), and the horizontal line within the box represents the median. Black-filled and open circles represent outliers (1.5 times beyond the IQR) and extreme outliers (3 times beyond the IQR), respectively. The whisker delineates the smallest and the largest non-outlier observations. **D**, Y chromosome phylogeny of the Bronze Age Xinjiang male individuals. Xiaohe male individuals fall into a branch distinct from western Bronze Age steppe pastoralists, such as Afanasievo and Yamnaya. One individual from Beifang falls in a position that is more basal than Xiaohe, but its phylogenetic position cannot be fixed due to low coverage, and its proximate position(s) are instead indicated with an asterisk.




**Extended Data Fig. 5 | Proteomic evidence for dairy consumption in Xiaohe dental calculus, ca. 2000-1800 BCE. A**, B-and Y-ion series for the frequently observed β-lactoglobulin peptide TPEVD(D/N/K)EALEK, which contains a taxon-specific polymorphic residue: D, Bovinae; N, *Ovis*; K, *Capra*. See SI Appendix. **B**, Taxonomically assigned β-lactoglobulin (black), α-S1-casein (dark grey), and α-lactalbumin peptide spectral matches (PSMs) presented as scaled pie charts on a cladogram of dairy livestock. Bracketed numbers represent the number of PSMs (excluding duplicates) assigned to each node. † Included on the Bovidae node are: 13 PSMs assigned to Bovidae; 21 PSMs assigned to Bovidae but excluding *Capra*.

## Extended Data Table 1 | A summary of the Bronze Age Xinjiang individuals reported in this study

| Sample ID | Group ID | Date (BCE) | Archaeol-ogical site | Skeletal elements | UDG treatment | SNPs | Mean coverage | Biological sex | mtDNA haplogroup | Y. chr. haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|
| AYIM22BY | Dzungaria_EBA1 | **2843-2811** | Ayituohan | Tooth | no | 495853 | 0.7010 | M | U5a1a1 | Q1b1 |
| AYIM22BN | Dzungaria_EBA1 | 2800-2600 | Ayituohan | Petrous bone | no | 524334 | 0.8990 | F | T2d1a | — |
| SSGM16 | Dzungaria_EBA1 | **2863-2801** | Songshugou | Tooth | no | 785354 | 1.4379 | F | H2b | — |
| G218M5-2 | Dzungaria_EBA2 | **2907-2851** | Nileke | Tooth | no | 790831 | 1.3523 | M | H15b1 | R1b1a1a2a2 |
| G218M5-3N | Dzungaria_EBA2 | **3005-2987** | Nileke | Tooth | no | 415959 | 0.5992 | M | U5a'b | Q1b1 |
| GMGM1 | Tarim_EMBA1 | **2135-2074** | Gumugou | Radius | no | 269355 | 0.3696 | F | C4 | — |
| XHM100 | Tarim_EMBA1 | **1884-1740** | Xiaohe | Tooth | half | 356623 | 0.4014 | F | C4 | — |
| XHM110 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 514292 | 0.6378 | F | C4 | — |
| XHM135 | Tarim_EMBA1 | **1936-1860** | Xiaohe | Tooth | half | 781191 | 0.6301 | F | C4 | — |
| XHM75 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 227235 | 0.2471 | F | C4 | — |
| L5209 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 767789 | 1.2160 | M | C4 | R1b1c |
| L5213 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 808182 | 1.2051 | M | R1b1 | R1b1c |
| L4964 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 60893 | 0.0632 | F | C4 | — |
| L6101 | Tarim_EMBA1 | **1767-1623** | Xiaohe | Tooth | half | 160220 | 0.1813 | F | C4 | — |
| L6103 | Tarim_EMBA1 | **1785-1664** | Xiaohe | Tooth | half | 536220 | 1.0811 | F | C4 | — |
| L6105 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 490934 | 0.5782 | F | C4 | — |
| L6106 | Tarim_EMBA1 | 2000-1800 | Xiaohe | Tooth | half | 247447 | 0.3530 | F | C4 | — |
| 11KBM1 | Tarim_EMBA2 | **1876-1839** | Beifang | Tooth | half | 131030 | 0.1943 | M | C4 | R1 (xR1a, xR1b1) |

In the "Date (BCE)" column, individuals directly dated by AMS are marked in bold (calibrated dates with 95.4% confidence interval) while the remaining dates are based on the archaeological contexts. The "SNPs" column shows the number of SNPs in the 1240k panel covered in each individual. Genome-wide data of seven individuals (L5209, L5213, L4964, L6101, L6103, L6105, L6106) were generated from MPP by enriching endogenous DNA for the 1240k panel SNPs.



## Extended Data Table 2 | Dietary proteins identified in the dental calculus of individuals analyzed from the Tarim Basin Xiaohe cemetery

| Sample ID | Date (BCE) | Archaeological layer | Milk Proteins | Total PSMs | Livestock |
|---|---|---|---|---|---|
| XHM100 | **1884-1740** | 5 | β-lactoglobulin | 50 | Cattle or sheep |
| XHM109 | 2000-1800 | 4 | β-lactoglobulin | 29 | Cattle or sheep |
| XHM112 | 2000-1800 | 5 | β-lactoglobulin | 23 | Cattle |
| XHM115 | 2000-1800 | 5 | β-lactoglobulin, α-S1-casein, α-lactalbumin | 98 | Cattle, goat, sheep |
| XHM117 | 2000-1800 | 5 | β-lactoglobulin, α-S1-casein, α-lactalbumin | 132 | Cattle or sheep |
| XHM125 | 2000-1800 | 4 | β-lactoglobulin | 41 | Cattle |
| XHM135 | **1936-1860** | 5 | β-lactoglobulin | 10 | Cattle or sheep |

In the "Date (BCE)" column, individuals directly dated by AMS are marked in bold (calibrated dates with 95.4% confidence interval) while the remaining dates are based on the archaeological contexts. The "Livestock" column shows consensus taxonomic assignment based on observed amino acid variants in milk peptides.

## Extended Data Table 3 | Robustness of key qpAdm admixture models

**A. No BMAC/IAMC-related ancestry component in Dzungaria_EBA**

| Target | Ref1 | Ref2 | Ref3 | Pval | Coef(Ref1) | Coef(Ref2) | Coef(Ref3) | Extra outgroups |
|---|---|---|---|---|---|---|---|---|
| Dzungaria_EBA1 | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 2.90E-01 | 0.717 ± 0.024 | 0.192 ± 0.044 | 0.091 ± 0.026 | |
| | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 3.06E-01 | 0.710 ± 0.022 | 0.212 ± 0.036 | 0.078 ± 0.020 | Geoksyur_EN |
| | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 3.61E-01 | 0.712 ± 0.022 | 0.205 ± 0.036 | 0.083 ± 0.020 | BMAC |
| | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 1.41E-01 | 0.754 ± 0.022 | 0.169 ± 0.026 | 0.078 ± 0.018 | Geoksyur_EN+AG3 |
| Dzungaria_EBA2 | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 1.48E-02 | 0.532 ± 0.030 | 0.359 ± 0.057 | 0.109 ± 0.034 | |
| | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 1.60E-02 | 0.545 ± 0.027 | 0.325 ± 0.044 | 0.130 ± 0.025 | Geoksyur_EN |
| | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 2.16E-02 | 0.540 ± 0.027 | 0.336 ± 0.044 | 0.125 ± 0.025 | BMAC |
| | Afanasievo | Tarim_EMBA1 | Baikal_EBA | 5.08E-01 | 0.527 ± 0.026 | 0.349 ± 0.033 | 0.125 ± 0.023 | Geoksyur_EN+AG3 |

**B. Comparison of Dzungaria_EBA1 and Afanasievo as a source for Chemurchek and IAMC populations**

| Target | Ref1 | Ref2 | Ref3 | Pval | Coef(Ref1) | Coef(Ref2) | Coef(Ref3) | Extra outgroups |
|---|---|---|---|---|---|---|---|---|
| Chemurchek_merged | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 4.92E-01 | 0.241 ± 0.044 | 0.673 ± 0.081 | 0.087 ± 0.046 | |
| | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 4.74E-01 | 0.260 ± 0.038 | 0.618 ± 0.057 | 0.122 ± 0.029 | Afanasievo |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 4.29E-03 | 0.466 ± 0.021 | 0.440 ± 0.045 | 0.094 ± 0.039 | |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 5.24E-04 | 0.454 ± 0.021 | 0.403 ± 0.044 | 0.143 ± 0.036 | Dzungaria_EBA1 |
| Dali_EBA | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 6.03E-01 | 0.537 ± 0.053 | 0.303 ± 0.096 | 0.159 ± 0.055 | |
| | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 7.05E-01 | 0.529 ± 0.049 | 0.326 ± 0.073 | 0.145 ± 0.036 | Afanasievo |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 7.58E-01 | 0.634 ± 0.029 | 0.248 ± 0.065 | 0.118 ± 0.055 | |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 8.18E-01 | 0.633 ± 0.029 | 0.245 ± 0.060 | 0.122 ± 0.047 | Dzungaria_EBA1 |
| Aigyrzhal_BA | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 1.69E-01 | 0.140 ± 0.042 | 0.377 ± 0.079 | 0.483 ± 0.045 | |
| | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 1.93E-01 | 0.160 ± 0.035 | 0.326 ± 0.054 | 0.514 ± 0.029 | Afanasievo |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 1.29E-02 | 0.278 ± 0.020 | 0.211 ± 0.050 | 0.512 ± 0.043 | |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 1.60E-02 | 0.273 ± 0.020 | 0.203 ± 0.046 | 0.525 ± 0.038 | Dzungaria_EBA1 |
| Kanai_MBA | Tarim_EMBA1 | Dzungaria_EBA1 | Baikal_EBA | 8.74E-01 | 0.324 ± 0.118 | 0.380 ± 0.076 | 0.296 ± 0.058 | |
| | Tarim_EMBA1 | Dzungaria_EBA1 | Baikal_EBA | 6.89E-01 | 0.211 ± 0.092 | 0.427 ± 0.071 | 0.363 ± 0.037 | Afanasievo |
| | Tarim_EMBA1 | Afanasievo | Baikal_EBA | 7.65E-01 | 0.316 ± 0.102 | 0.316 ± 0.054 | 0.368 ± 0.060 | |
| | Tarim_EMBA1 | Afanasievo | Baikal_EBA | 8.93E-01 | 0.371 ± 0.075 | 0.293 ± 0.047 | 0.336 ± 0.043 | Dzungaria_EBA1 |
| Kumsay_EBA | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 5.95E-04 | 0.317 ± 0.043 | 0.561 ± 0.079 | 0.122 ± 0.043 | |
| | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 1.18E-03 | 0.304 ± 0.036 | 0.592 ± 0.054 | 0.104 ± 0.027 | Afanasievo |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 7.97E-02 | 0.488 ± 0.018 | 0.416 ± 0.041 | 0.097 ± 0.034 | |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 9.14E-02 | 0.488 ± 0.018 | 0.388 ± 0.040 | 0.124 ± 0.031 | Dzungaria_EBA1 |
| Mereke_MBA | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 1.60E-01 | 0.195 ± 0.048 | 0.810 ± 0.087 | -0.005 ± 0.048 | |
| | Tarim_EMBA1 | Dzungaria_EBA1 | Geoksuyr_EN | 9.55E-02 | 0.229 ± 0.039 | 0.718 ± 0.058 | 0.052 ± 0.028 | Afanasievo |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 1.84E-01 | 0.455 ± 0.021 | 0.566 ± 0.047 | -0.021 ± 0.039 | |
| | Tarim_EMBA1 | Afanasievo | Geoksuyr_EN | 2.29E-03 | 0.443 ± 0.021 | 0.503 ± 0.044 | 0.055 ± 0.034 | Dzungaria_EBA1 |

**C. Modeling of ANE-rich pre-Bronze Age populations in Central Asia**

| Target | Ref1 | Ref2 | Ref3 | Pval | Coef(Ref1) | Coef(Ref2) | Coef(Ref3) | Extra outgroups |
|---|---|---|---|---|---|---|---|---|
| Botai_CA | Tarim_EMBA1 | Baikal_EBA | EHG | 2.57E-01 | 0.695 ± 0.057 | 0.056 ± 0.029 | 0.249 ± 0.034 | |
| | Tarim_EMBA1 | Baikal_EBA | EHG | 1.39E-01 | 0.554 ± 0.036 | 0.133 ± 0.019 | 0.312 ± 0.028 | AG3 |
| | Tarim_EMBA1 | Baikal_EBA | EHG | 1.90E-02 | 0.581 ± 0.039 | 0.121 ± 0.017 | 0.298 ± 0.029 | Afanasievo |
| | Tarim_EMBA1 | Baikal_EBA | EHG | 1.78E-01 | 0.639 ± 0.039 | 0.084 ± 0.020 | 0.277 ± 0.027 | Geoksyur_EN |
| West_Siberia_N | Tarim_EMBA1 | - | EHG | 3.92E-01 | 0.671 ± 0.030 | - | 0.329 ± 0.030 | |
| | Tarim_EMBA1 | - | EHG | 2.22E-01 | 0.674 ± 0.033 | - | 0.326 ± 0.033 | AG3 |
| | Tarim_EMBA1 | - | EHG | 2.81E-01 | 0.669 ± 0.030 | - | 0.331 ± 0.030 | Afanasievo |
| | Tarim_EMBA1 | - | EHG | 4.10E-01 | 0.680 ± 0.029 | - | 0.320 ± 0.029 | Geoksyur_EN |

We present details of key qpAdm admixture models reported in this study with alternative outgroup sets including sources from competing admixture models. "Coef" columns show the ancestry proportion and its standard error, calculated by 5 cM block jackknifing. Extra outgroup shows outgroups added to the base set. (A) Admixture models for Dzungaria_EBA do not change when BMAC/IAMC-related populations are added to the outgroup, supporting no contribution from them. (B) For Chemurchek and IAMC populations, we compare models including Dzungaria_EBA1 or Afanasievo as a competing source. Dzungaria_EBA1 works better for Chemurchek, Aigyrzhal_BA, Mereke_MBA, while Afanasievo works better for Kumsay_EBA. (C) Admixture models for Botai_CA and West_Siberia_N robustly hold when Afanasievo, Geoksyur_EN, or AG3 are included as an additional outgroup. Pval represents qpAdm *p*-value for the one-sided likelihood ratio test comparing the nested model (i.e. the target population is a mixture of the given references) with the nesting one (i.e. the target population cannot be sufficiently modeled as mixture of the given references). *P*-values are not multiple-testing corrected. Standard error measures were calculated with 5 cM block jackknifing.

# 6. Manuscript C

**Manucript Nr. : 3 (Manuscript C)**

**Title of the Manuscript:** Emergence and intensification of dairying in the Caucasus and Eurasian steppes

**Authors:** Ashley Scott, Sabine Reinhold, Taylor Hermes, Alexey A. Kalmykov, Andrey Belinskiy, Alexandra Buzhilova, Natalia Berezina, Anatoliy R. Kantorovich, Vladimir E. Maslov, Farhad Guliyev, Bertille Lyonnet, Parviz Gasimov, Tufan Axhundov, Bakhtiyar Jalilov, Jeyhun Eminli, Emil Iskandarov, Emily Hammer, Selin Nugent, Richard Hagan, Kerttu Majander, Päivi Onkamo, Kerkko Nordqvist, Natalia Shishlina, Elena Kaverzneva, Arkadiy I. Korolev, Aleksandr A. Khokhlov, Roman V. Smolyaninov, Rüdiger Krause, Eliza Stolarzyk, Maria Karapetian, Svetlana V. Sharapova, Johannes Krause, Svend Hansen, Wolfgang Haak, Christina Warinner

**Bibliographic Information:** Scott et al. (submitted for publication)

**The candidate is:** ☒ First author, ☐ Co-first author, ☐ Corresponding author, ☐ coauthor

**Status:** submitted for publication, *Nature Evolution and Ecology*, 16.09. 2021.

**Proportion (in %) of authors in the publication:** (to be indicated from 20%)

| Author | Concept | Data analysis | Experimental | Manuscript composition | Material provision |
|---|---|---|---|---|---|
| Scott A | | 100% | 90% | 35% | |
| Reinhold S | | | | 20% | |
| Hermes T | | | | | |
| Kalmykov AA | | | | | |
| Belinskiy A | | | | | |
| Buzhilova A | | | | | |
| Berezina N | | | | | |
| Kantorovich AR | | | | | |
| Maslov VE | | | | | |
| Guliyev F | | | | | |
| Lyonnet B | | | | | |
| Gasimov P | | | | | |
| Axhundov T | | | | | |
| Jalilov B | | | | | |
| Eminli J | | | | | |
| Iskandarov E | | | | | |
| Hammer E | 20% | | | | |
| Nugent S | | | | | |
| Hagan R | | | 10% | | |
| Majander, K | | | | | |
| Onkamo P | | | | | |
| Nordqvist K | | | | | |
| Shishlina N | | | | | |
| Kaverzneva E | | | | | |
| Korolev AI | | | | | |
| Khokhlov AA | | | | | |
| Smolyaninov RV | | | | | |
| Krause R | | | | | |
| StolarzykE | | | | | |
| Karapetian M | | | | | |
| Sharapova S | | | | | |
| Krause J | 20% | | | | |
| Hansen S | 20% | | | | |
| Haak W | 20% | | | | |
| Warinner C | 20% | | | 35% | |

nature ecology & evolution

Articles

https://doi.org/10.1038/s41559-022-01701-6

OPEN

# Emergence and intensification of dairying in the Caucasus and Eurasian steppes

Ashley Scott[1,2,3], Sabine Reinhold[4], Taylor Hermes[1,2], Alexey A. Kalmykov[5], Andrey Belinskiy[5], Alexandra Buzhilova[6], Natalia Berezina[6], Anatoliy R. Kantorovich[7], Vladimir E. Maslov[8], Farhad Guliyev[9], Bertille Lyonnet[10], Parviz Gasimov[9], Bakhtiyar Jalilov[9], Jeyhun Eminli[9], Emil Iskandarov[9], Emily Hammer[11], Selin E. Nugent[12], Richard Hagan[1,13], Kerttu Majander[1,14], Päivi Onkamo[15,16], Kerkko Nordqvist[17], Natalia Shishlina[18,19], Elena Kaverzneva[18], Arkadiy I. Korolev[20], Aleksandr A. Khokhlov[20], Roman V. Smolyaninov[21], Svetlana V. Sharapova[22], Rüdiger Krause[23], Marina Karapetian[6], Eliza Stolarczyk[23], Johannes Krause[1,2], Svend Hansen[3] ✉, Wolfgang Haak[1,2] ✉ and Christina Warinner[1,2,24] ✉

Archaeological and archaeogenetic evidence points to the Pontic–Caspian steppe zone between the Caucasus and the Black Sea as the crucible from which the earliest steppe pastoralist societies arose and spread, ultimately influencing populations from Europe to Inner Asia. However, little is known about their economic foundations and the factors that may have contributed to their extensive mobility. Here, we investigate dietary proteins within the dental calculus proteomes of 45 individuals spanning the Neolithic to Greco-Roman periods in the Pontic–Caspian Steppe and neighbouring South Caucasus, Oka–Volga–Don and East Urals regions. We find that sheep dairying accompanies the earliest forms of Eneolithic pastoralism in the North Caucasus. During the fourth millennium bc, Maykop and early Yamnaya populations also focused dairying exclusively on sheep while reserving cattle for traction and other purposes. We observe a breakdown in livestock specialization and an economic diversification of dairy herds coinciding with aridification during the subsequent late Yamnaya and North Caucasus Culture phases, followed by severe climate deterioration during the Catacomb and Lola periods. The need for additional pastures to support these herds may have driven the heightened mobility of the Middle and Late Bronze Age periods. Following a hiatus of more than 500 years, the North Caucasian steppe was repopulated by Early Iron Age societies with a broad mobile dairy economy, including a new focus on horse milking.

During the early to mid-Holocene (ca. 9.0–3.5 thousand years ago (kya)), dairying played a vital role in the development of human food systems across Europe, Africa and Asia[1–8]. Early agropastoral societies raised livestock animals that could provide them with milk, meat, wool, leather and traction[9], and milk rose to prominence as an especially important, nutrient-rich food source. Milk is rich in protein, fat, sugar (lactose), vitamins and minerals, such as calcium[10], and the water content in milk can be relied on in times of drought or scarcity[11,12]. Although milk itself is highly perishable, it can be transformed through microbial fermentation and other forms of manipulation into more stable products, such as yogurt, butter, ghee, cheese and curds, that can be stored for longer periods in surplus[13–15].

First attested in Anatolia during the seventh and sixth millennia bc[3,6], ruminant dairying subsequently spread to both Europe and Africa by the late sixth millennium bc[4,16], but less is known about its initial dispersals into Asia[17–19]. One major vector by which dairying spread was the Eurasian steppe, an enormous expanse of grasslands stretching 6,000 km from the Carpathian Basin to Mongolia. Recent studies have traced the introduction of dairying in Mongolia

[1]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany. [3]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilian University Munich, Munich, Germany. [4]Eurasia Department, German Archaeological Institute, Berlin, Germany. [5]'Nasledie' Cultural Heritage Unit, Stavropol, Russia. [6]Research Institute and Museum of Anthropology, Lomonosov Moscow State University, Moscow, Russia. [7]Department of Archaeology, Faculty of History, Lomonosov Moscow State University, Moscow, Russia. [8]Institute of Archaeology RAS, Moscow, Russian Federation. [9]Department of Humanitarian and Social Sciences, Institute of Archaeology, Ethnography and Anthropology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan. [10]PROCLAC/UMR 7192 Laboratory, French National Centre for Scientific Research, Paris, France. [11]Near Eastern Languages and Civilizations and Price Lab for the Digital Humanities, University of Pennsylvania, Philadelphia, PA, USA. [12]Faculty of Technology, Design & Environment, Oxford Brookes University, Oxford, UK. [13]Department of Archaeology, University of York, York, UK. [14]Institute of Evolutionary Medicine, University of Zürich, Zürich, Switzerland. [15]Department of Biology, University of Turku, Turku, Finland. [16]Department of Biosciences, University of Helsinki, Helsinki, Finland. [17]Department of Cultures, University of Helsinki, Helsinki, Finland. [18]State Historical Museum, Moscow, Russia. [19]Peter the Great Museum of Anthropology and Ethnography (the Kunstkamera), Saint Petersburg, Russia. [20]Department of History and Archaeology, Samara State University of Social Sciences and Education, Samara, Russia. [21]Lipetsk State Pedagogical University, Lipetsk, Russia. [22]Institute of History and Archaeology, Ural Branch of the Russian Academy of Science, Ekaterinburg, Russia. [23]Department of Archaeological Sciences, Johann Wolfgang Goethe University, Frankfurt am Main, Germany. [24]Department of Anthropology, Harvard University, Cambridge, MA, USA. ✉e-mail: svend.hansen@dainst.de; wolfgang_haak@eva.mpg.de; warinner@fas.harvard.edu



Content courtesy of Springer Nature, terms of use apply. Rights reserved

to ca. 3000 bc with the appearance of mobile steppe herders associated with the Early Bronze Age Afanasievo culture[2], a group with close genetic and cultural ties to pastoralists on the Pontic–Caspian steppe, most notably the Yamnaya culture (ca. 3300–2500 bc)[20–23]. Populations from the Pontic–Caspian steppe are also linked to Late Neolithic and Bronze Age westward expansions, including the emergence of the Corded Ware (2900–2200 bc) and Bell Beaker (2750–1800 bc) phenomena in Europe[24–27]. Understanding the population and economic history of the Pontic–Caspian steppe, the source region for these continental-scale expansions during the third millennium bc, is critical for revealing the main factors that drove the heightened mobility of Eneolithic and Early Bronze Age pastoralists in Eurasia.

When Pontic–Caspian steppe populations first began dairying and how their animal management strategies may have influenced their mobility and subsequent migrations remain poorly known. From the Mesolithic through the Eneolithic, populations living in the southern Russian plain and Caucasus region primarily hunted local wild game, which included aurochs (*Bos primigenius*), saiga antelope (*Saiga tatarica*), red deer (*Cervus elaphus*), tarpan (*Equus ferus*), onager (*Equus hemionus*) and wild boar (*Sus scrofa*), as well as birds, fish and molluscs[28–31]. Animal husbandry of domesticated sheep (*Ovis aries*), goats (*Capra hircus*), cattle (*Bos taurus*) and pigs (*Sus scrofa*) spread to the North Caucasian steppe from Anatolia during the fifth millennium bc by either a circum-Pontic route[28] or by crossing the Caucasus mountains from the south[32–35]. By the mid-fifth millennium bc, agropastoralists of the Cucuteni–Trypillia culture in Ukraine were regularly interacting with steppe populations north of the Black Sea[36], and Eneolithic populations genetically related to South Caucasian and Anatolian agropastoralist groups had become established in the North Caucasus piedmont steppe[32,33,37] and were part of a broader Mesopotamian interaction sphere[38,39].

After the introduction of animal husbandry to the region, Bronze Age steppe populations innovated a new economic system of mobile pastoralism focused on sheep and cattle[40], and settlements became effectively absent on the steppe for the next two millennia[40,41]. This new, more mobile form of pastoralism is first evident among Steppe Late Maykop groups (3500–2900 bc), who fall broadly within the Late Maykop cultural sphere but are genetically distinct from their higher-elevation counterparts[33], and fully mobile pastoralism subsequently became the predominant subsistence strategy on the steppe with the Yamnaya culture (3,300–2,500 bc)[41]. Horse domestication occurred during the third millennium on the Pontic–Caspian steppe[42,43], and, by the late third and early second millennium bc, domestic horses were increasingly part of the steppe mobile pastoralist economy[44] and had even spread to Anatolia and Mesopotamia through Pontic–Caspian–Transcaucasian interaction networks[45]. Mobile pastoralism continued among the Catacomb (2800–2200 bc) and North Caucasus Culture (NCC; 2800–2400 bc) groups in the steppe until worsening climatic conditions and aridification ca. 2300–2200 bc, in association with the 4.2 kya climate event[46,47], ultimately led to an abandonment of the steppe region by 1700 bc[40,41]. Despite their cultural differences, recent palaeogenomic analysis has shown that these Bronze Age steppe populations were genetically highly similar[33], which may, in part, reflect their mobile lifestyles and persistent multicultural interactions over millennia[40].

Throughout the Pontic–Caspian steppe, sheep, goat and cattle dominate most studied steppe archaeofaunal collections from the fourth to second millennia bc[41,48,49]. Wheeled transport in the form of wagons first appears in kurgans (burial mounds) of the Steppe Late Maykop in the second half of the fourth millennium bc[50], and such technology is argued to be essential for enabling the household mobility required for mobile pastoralism[40]. Oxen teams dated to the same period and, later, horses and chariots in the second millennium bc, further facilitated mobility[51]. Sheep wool was present in the North Caucasus by the early third millennium bc, possibly having originated in Anatolia, and the use of wool subsequently spread across the steppe and into Inner Asia during the second millennium bc[52]. Among the region's major secondary products, dairying is argued to have possibly emerged first[50], in part because dairying was already well established in both Anatolia and surrounding regions by the sixth millennium bc[6,53–55], whereas evidence for traction and wool are only attested millennia later. Nevertheless, current evidence for early dairying in the Pontic–Caspian steppe is, until now, only attested on its eastern fringes[7]. Previous isotopic studies have been unable to identify clear indications of dairy consumption, finding instead non-specific evidence for high consumption of animal protein and a highly complex isoscape, reflecting both ecological diversity and temporal climatic shifts[41,48,56]. However, the isotopic data suggest a stronger contribution of sheep or goat products to the human diet than those from cattle[41]. Few zooarchaeological studies have systematically investigated herd management and mortality profiles in the region, but the earliest agropastoralist communities in the North Caucasus piedmont steppe were not thought to have engaged in dairying[49]. Likewise, there are few indications of animal management for milk production among Neolithic agropastoralist communities in the South Caucasus[42]. Rather, it is only in the second millennium bc that zooarchaeological studies from Late Bronze Age settlements in the Caucasus have found clear evidence for the deliberate keeping of sheep for milk production[57,58], and it is only later during the Iron Age that cattle show mortality profiles consistent with dairying[59].

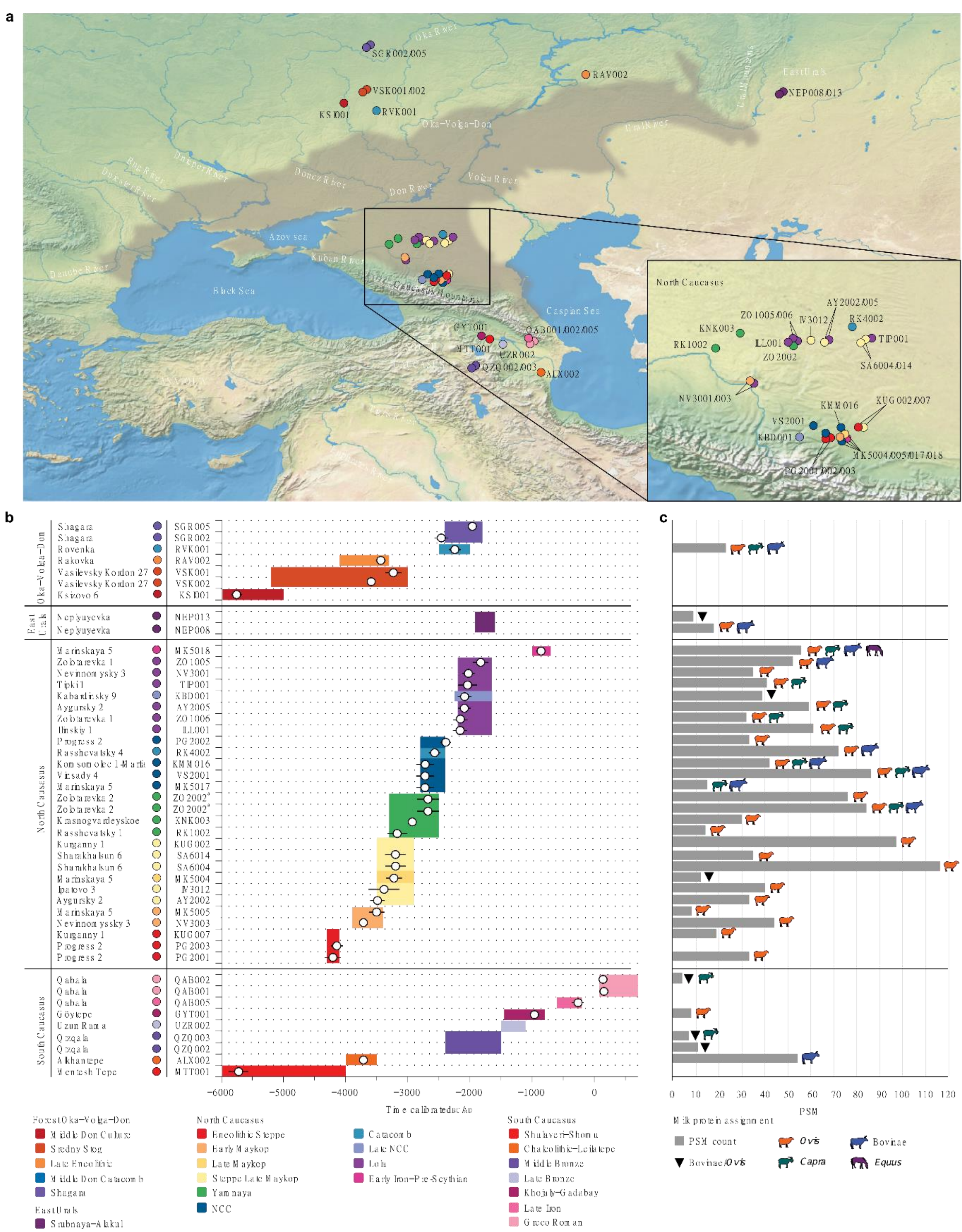
The absence of settlements on the steppe and the near-exclusive archaeological focus on mortuary contexts have made it difficult to reconstruct the nature and extent of dairying in the wider North Caucasian pastoralist economy. In this article, we apply high-resolution tandem mass spectrometry to human dental calculus from 45 individuals at 29 sites in the North Caucasus ($n$ = 27) and the neighbouring South Caucasus ($n$ = 9), Oka–Volga–Don ($n$ = 7) and East Urals ($n$ = 2) regions (Fig. 1a,b, Supplementary Data 1 and Supplementary Information) to identify evidence of consumed dairy proteins in populations spanning the Neolithic to the Greco-Roman periods (ca. 6000 bc to 200 ad). We find that dairy products were consumed in the North Caucasus from the late fifth millennium bc onwards and that a dairy-inclusive subsistence characterizes even the Eneolithic populations in the piedmont and steppe zones. Dairy consumption was prevalent for all analysed periods and ecotones in the North Caucasus, with milk proteins identified in 26 of 27 tested individuals. We identify an initial, near-exclusive dairying focus on sheep among the Maykop, Steppe Maykop and early Yamnaya, followed by diversification within the late Yamnaya, NCC and Catacomb cultures during the Middle Bronze Age to additionally incorporate goat and cattle milking.

Fig. 1 | **Map and timeline of sites and individuals in the study and milk protein results.** a, Map of study area and major cultural regions mentioned in the text: Oka–Volga–Don, East Urals, North Caucasus, South Caucasus and Anatolia. Extent of the Pontic–Caspian steppe is shown in grey. Inset: enhanced view of North Caucasus sites. b, Timeline of sites and individuals analysed in this study. Individuals are organized by region, with archaeological culture or period indicated by colour corresponding to the legend. White circles indicate median calibrated radiocarbon dates, and error bars are 2 s.d. Coloured bars display the time spans conventionally associated with the archaeological cultures and time periods. c, Milk protein evidence by individual, displayed as total PSM count to the milk proteins BLG, alpha-lactalbumin and alpha-S1-casein. Consensus livestock assignment was determined by parsimony. [a]Two dental calculus samples were analysed from ZO 2002. Basemap is from https://www.naturalearthdata.com/.



Content courtesy of Springer Nature, terms of use apply. Rights reserved

Later, during the Early Iron Age, we observe direct evidence of horse milk consumption in association with pre-Scythian groups repopulating the steppe after a centuries-long hiatus. In the South Caucasus, we identify evidence of cattle milking (ca. 3700 bc) nearly 1,000 years before we first observe it in the North Caucasus (ca. 2700 bc), and, in the Oka–Volga–Don region, we observe





Content courtesy of Springer Nature, terms of use apply. Rights reserved

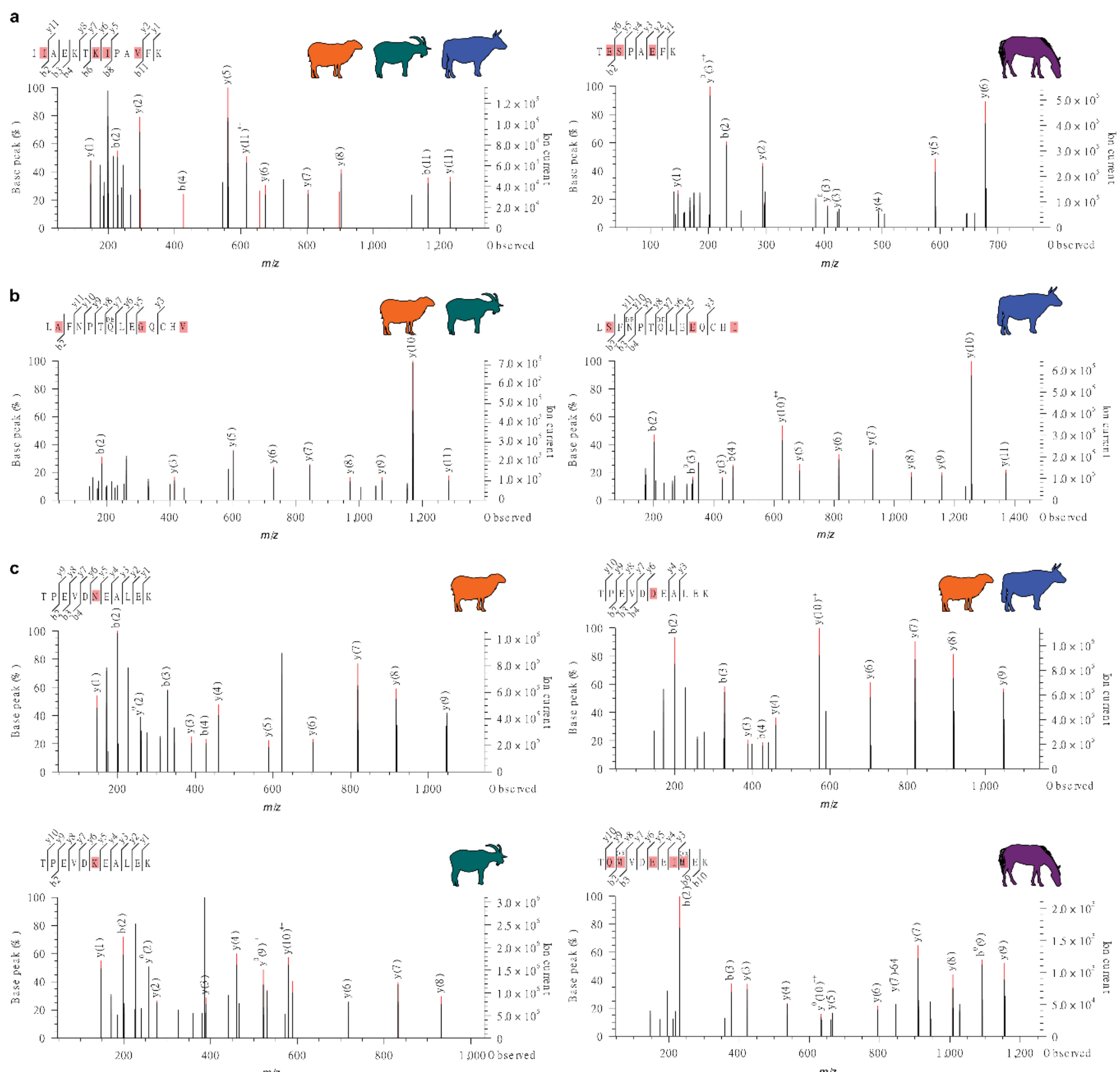



Fig. 2 | **Representative tandem mass spectrometry spectra of selected BLG peptides with differing levels of taxonomic resolution observed in this study.** a, Overall, most BLG sequences were highly conserved among bovids (left) but distinct from equids (right). Spectra originate from AY2005 and MK5018. b, Among bovids, the BLG C-terminus peptide distinguishes caprines (left) and bovines (right). Spectra originate from VS2001 and VS2001. c, The most frequently observed peptide reliably distinguishes *Ovis* (upper left), *Capra* (lower left) and *Equus* (lower right) but cannot distinguish *Ovis* and Bovinae due to the ambiguity of the sixth residue, which may be aspartic acid (Bovinae) or deamidated asparagine (*Ovis*)[5] (upper right). Spectra originate from KUG007, RK4002, VS2001 and MK5018. The b- and y-ion series is shown at the top left of each spectrum, and taxonomically informative residues within the peptide sequence are highlighted in pink. A comprehensive list of all identified PSMs and taxonomic assignments is provided in Supplementary Data 3.

limited evidence of dairying, beginning only during the second millennium BC.

## Results

Milk proteins were identified in 34 of 45 analysed individuals across all time periods (Fig. 1c and Supplementary Data 1). Protein recovery in 31 individuals was sufficient to allow the identification of major ruminant livestock milks from sheep (*Ovis*), goat (*Capra*) and/or cattle (*Bos*/Bovinae), whereas the milk proteins of three individuals were represented by non-specific bovid peptides, indicating either sheep or cattle. Additionally, one individual had taxonomically distinctive peptide spectral matches (PSMs) to *Equus* milk proteins. Beta-lactoglobulin (BLG), which was detected for all dairy livestock (Fig. 2), was the most prevalent and abundant milk protein detected, a pattern consistent with previous studies of dental calculus[2,6,60]. In addition to BLG, which was identified in all 34 milk-positive individuals, we also identified the whey protein alpha-lactalbumin in two individuals and the



Content courtesy of Springer Nature, terms of use apply. Rights reserved

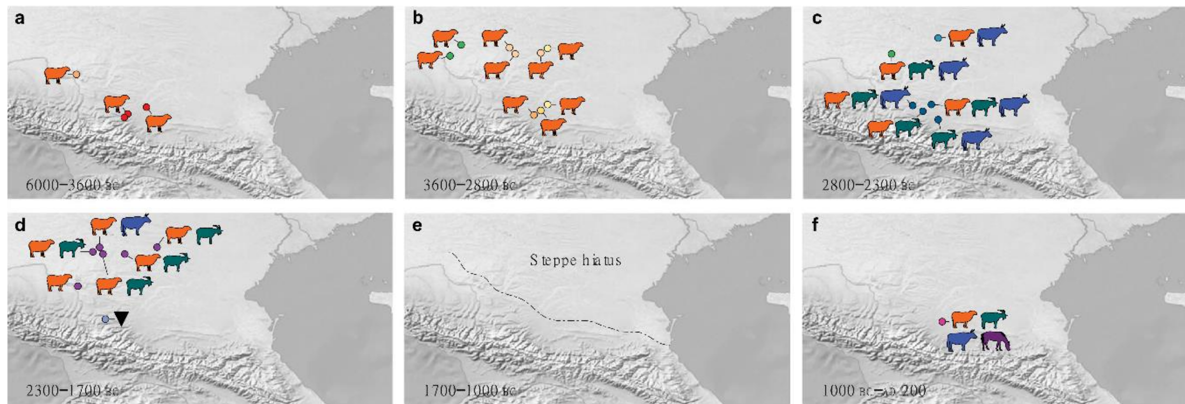



Fig. 3 | **Changing dairy patterns through time in the North Caucasus region.** a, During the Eneolithic and initial Bronze Age, dairying focused on sheep in the North Caucasus from 4200 bc onwards. b, Sheep dairying continued during the Early Bronze Age among the Maykop, Steppe Maykop and early Yamnaya. c, After 2800 bc, goat and cattle dairying appeared for the first time in the steppe, and diversified dairy economies of sheep, goats and cattle characterize the late Yamnaya, NCC and Catacomb cultures. d, Diversified dairy economies persisted among the post-Catacomb and Lola cultures, but with an increased focus on sheep and goats as environmental conditions declined. e, During the Late Bronze Age, the North Caucasus steppe was largely depopulated, and ca. 1700 bc a centuries-long hiatus began that corresponded to a period of extreme aridity. Dashed line shows the southern extent of depopulation. f, After 1000 bc, post-hiatus groups repopulated the steppe in the Early Iron Age, resuming sheep, goat and cattle milking and also introducing horse milking. Site colours and animal symbols correspond to those in Fig. 1. All tested individuals in the map extent are shown, including those without evidence of milk protein.

curd protein alpha-S1-casein in two individuals. All dental calculus samples yielded proteomes consistent with an oral microbiome profile, and age-associated N/Q deamidation was a top modification across the dataset (Supplementary Data 2). Milk protein peptide-level deamidation is reported in Supplementary Data 3. No dietary proteins were detected in non-template extraction controls (Supplementary Data 3).

**North Caucasus.** North of the Caucasus mountains, within a geographically and culturally contiguous region that encompasses the piedmont zone, steppe and southern Russian plain, we analysed dietary proteins within the dental calculus proteomes of 27 individuals, including three Eneolithic, 23 Bronze Age and one Early Iron Age individual. Overall, we identified milk proteins in 96% of individuals (*n* = 26) (Fig. 1c) and observed high levels of milk protein PSMs per individual (mean 47 ± 27; Supplementary Data 3), with milk peptides often being among the most abundant peptides identified in the dental calculus proteomes. Among Eneolithic individuals, two of three were positive for milk proteins. The oldest individual from this region in our study, PG2001 from the piedmont site of Progress 2 and dated to 4338–4074 bc, indicates that dairying has been a feature of the region's economy since at least the late fifth millennium bc. During the fourth and third millennia bc, we observed a continued reliance on dairying among all analysed Maykop and Steppe Maykop individuals (ca. 3900–2900 bc; *n* = 7), both in the piedmont and steppe zones as well as in all Yamnaya individuals (ca. 3300–2500 bc; *n* = 3). Notably, we detected only *Ovis* milk proteins at Eneolithic, Early Maykop, Late Maykop, Steppe Late Maykop and early Yamnaya sites, suggesting that dairying was a specialized activity focused on sheep during the fourth and fifth millennia bc (Fig. 3a,b). At the start of the early third millennium bc, we identified a broad shift in pastoralist practices towards more diversified dairying based on sheep, goat and cattle milk (Fig. 3c,d). Milk proteins from these three ruminant species were identified among individuals associated with the late Yamnaya (ca. 2850–2500 bc; *n* = 1), NCC (ca. 2800–2400 bc; *n* = 4), Catacomb (ca. 2800–2400 bc; *n* = 1), late NCC (ca. 2200–1650 bc, *n* = 1) and Lola/post-Catacomb (ca. 2200–1650 bc; *n* = 6) cultures, with most individuals having consumed the dairy products of two or three different animal milks in the form of sheep and goat milk, sheep and cattle milk or sheep, goat and cattle milk (Fig. 1 and Supplementary Data 3). Finally, during the Early Iron Age (eighth–fifth centuries bc), we observed the incorporation of horse (*Equus*) milk into the dairy economy (Fig. 3e), with *Ovis*, *Capra*, *Bos* and *Equus* milk proteins identified in the dental calculus of individual MK5018.

**South Caucasus.** In the South Caucasus, we analysed dietary proteins within the dental calculus proteomes of nine individuals dating from the Neolithic to Greco-Roman periods and identified milk proteins in half of the analysed individuals (Fig. 1c and Supplementary Data 3). No milk proteins were detected in the earliest individual, MTT001, dated to 5879–5562 bc from the Neolithic site of Mentesh Tepe associated with the Shomutepe–Shulaveri Culture. However, milk proteins were detected from the fourth millennium bc onwards in individuals dating to the Chalcolithic at Alkhantepe (*n* = 1), the Middle Bronze Age at Qızqala (*n* = 2), the Iron Age at Göytepe (*n* = 1) and the Greco-Roman period at Qabala (*n* = 1). Unlike in the North Caucasus, we did not observe an early focus on sheep dairying; rather, the earliest detected milk protein, identified in individual ALX002 dating to 3776–3651 bc, was assigned to cattle (Bovinae). Overall, we identified cattle (Bovinae), goat (*Capra*) and sheep (*Ovis*) milk protein in the South Caucasus but no horse (*Equus*) milk in any period there (Fig. 3).

**Oka–Volga–Don region.** In the Oka–Volga–Don region, we analysed dietary proteins within the dental calculus proteomes of seven individuals dating from the Eneolithic through the Middle Bronze Age (Fig. 1c and Supplementary Data 3). Despite excellent protein recovery, no milk proteins were detected in an individual from the Neolithic–Bronze Age site of Ksizovo 6, dating to 5837–5670 bc, nor from individuals associated with the Sredny Stog culture (*n* = 2) at the Eneolithic–Bronze Age site of Vasilevsky Kordon 27, dating to ca. 3600–3100 bc. Milk proteins were also absent from individual RAV002, dating to 3514–3356 bc, and from two Middle Bronze



Content courtesy of Springer Nature, terms of use apply. Rights reserved

Age individuals from the Shagara cemetery, dating to 2572–1893 bc. Only an individual at the site of Rovenka tested positive for milk proteins. This individual, RVK001, was associated with a late Catacomb culture site, dating to 2339–2148 bc, and was positive for sheep (*Ovis*), goat (*Capra*) and cattle (Bovinae) milk proteins (Fig. 1c).

**East Urals region.** We analysed two individuals from the East Urals region at the site of Neplyuyevka associated with the Late Bronze Age Srubnaya–Alakul cultural variant and dating to ca. 1900–1600 bc (Fig. 1c). We detected milk proteins for both individuals, identifying sheep (*Ovis*) and cattle (Bovinae) peptide sequences for NEP008 and non-specific bovid peptide sequences indicating either sheep or cattle for NEP013 (Supplementary Data 3).

## Discussion

**Eneolithic populations practiced dairy pastoralism.** Our results provide robust evidence that sheep dairying was practiced among fifth millennium bc Eneolithic groups in the North Caucasus piedmont and steppe zones. This finding resolves long-standing questions about the antiquity of dairying in the North Caucasus[47], as well as the species focus of early dairy herds, and it contributes to a growing body of evidence that dairying was likely introduced with domesticated livestock into the North Caucasus from the south. Recent palaeogenomic studies identified a genetic cline connecting Neolithic populations in eastern Anatolia and the South Caucasus that likely formed as early as 6500 bc[32], and continued population interaction into the Chalcolithic and Early Bronze Age periods (5500–3000 bc) suggests that these regions maintained close contact, with animal husbandry focused on pigs and ruminants also spreading via this corridor[61,62]. Early agropastoralists living in the northern Caucasus foothills associated with the Darkveti–Meshoko Eneolithic culture (ca. 4500–4000 bc) have a clear genetic connection to populations south of the Caucasus exhibiting Anatolian ancestry[33], suggesting a trans-Caucasian population expansion.

Although it has been speculated that dairying may have spread to the North Caucasus via these southern connections[50], few systematic zooarchaeological studies have been conducted, and the Eneolithic/Chalcolithic layers at the piedmont site of Meshoko Cave, which are among the best studied for the period[49], have yielded limited faunal remains, primarily of pigs, sheep, goats and cattle slaughtered at various ages. Subsequent attempts to clarify the agropastoralist economy using stable isotope analysis[41,48] have yielded equivocal results as to whether dairying was an Eneolithic or Bronze Age innovation in the North Caucasus. Here, through the identification of taxonomically informative peptides from the milk-specific protein BLG, we confirmed sheep milk consumption by Eneolithic individuals at the sites of Progress 2 and Kurganny 1. Notably, we found that dairy consumption was evident among individuals lacking Anatolian ancestry, such as PG2001[33], demonstrating that the adoption of dairying by North Caucasian transitional foragers was already underway during the late fifth millennium bc, which precedes Yamnaya expansions by a millennium.

**Maykop and Steppe Maykop dairy focused on sheep not cattle.** With the start of the fourth millennium bc, we found a continued reliance on dairy pastoralism revealed by ubiquitous evidence of milk consumption among all tested Maykop and Steppe Maykop individuals, further clarifying the high dependence of these groups on animal products[41,47]. Surprisingly, however, the dairy economy retained an apparent focus on sheep. Although sheep are known to have been important livestock for these groups[40,47], cattle feature more prominently at Maykop mortuary sites. They are modelled into gold and silver figurines[63], and an emphasis on the power and mobility of cattle is visible in funerary offerings of cattle crania, yokes and nose rings representing oxen teams[50]. Cattle also appear in bone assemblages at Maykop settlements[49]. The perishability of the major sheep secondary products of milk and wool, in contrast to the high visibility of cattle-associated material culture and skeletal remains, may have contributed to a biased understanding of the relative importance and roles of these livestock at Early Bronze Age sites[64]. Our results suggest that cattle were not important dairy livestock during this period and that there was probably a sharp division in livestock use among the Maykop and Steppe Maykop groups[41], with sheep being the primary targets of dairying and cattle mainly being used for traction and as a signifier of social identity and status.

**Dairy livestock diversified during Middle Bronze Age.** A change in dairying strategy to focus on more livestock species coincides with the Yamnaya horizon. Following the Maykop period, mobility expanded ever further with Yamnaya groups, who became the first permanently mobile pastoralists[17,44,65,66]. Although two early Yamnaya individuals analysed here yielded evidence of only sheep milk product consumption, a more diversified profile comprising sheep, goat and cattle milk was observed for a late Yamnaya individual at the site of Zolotarevka (ZO2002). This trend towards reliance on a broader range of dairy livestock continued and intensified during the Middle Bronze Age, when we observed a general diversification of pastoralist diets to include sheep, goat and cattle milk routinely. Most individuals of the Middle Bronze Age Catacomb, NCC, Late NCC and Lola cultures tested in this study consumed the dairy products of two or three livestock species. Palaeoecological studies have indicated that climate began to shift during the late Yamnaya phase, which also coincided with the first appearance of the Catacomb and NCC groups[48]. Before this, the climate experienced by the Maykop, Steppe Maykop and early Yamnaya was more favourable[67,68] and conducive to regular, short-distance annual mobility[47,48]. Subsequent aridification encouraged increased mobility, resulting in the exploitation of a wider range of steppe environments beyond the traditional Yamnaya cultural sphere to support livestock herds[40,48]. The shift to more diverse dairy herds in the North Caucasus also overlaps in time with Yamnaya expansions into southeastern Europe, as well as the parallel rise and expansion of the Corded Ware complex across northeastern and central Europe[27], suggesting that these events may be related to broader changes occurring within steppe and forest–steppe pastoralist societies at the time. Our results suggest that an initial diversification of production strategies to include sheep, goat and cattle milk may have functioned as an adaptation to an increasingly turbulent ecological setting, but this subsequently led to overgrazing and lasting damage to pastures due to ground compaction, soil nutrient loss and decreasing plant biomass[48,69]. At the end of the third millennium bc, coinciding with the emergence of the Lola culture, an intensified drought caused deflation and salinization of the soils in the already dwindling regional watersheds[40,69]. During the Lola period, water-demanding cattle may have decreased in dairying importance from the preceding Catacomb and NCC periods, as only one of six Lola individuals yielded evidence for cattle milk consumption. After 1700 bc, the steppe and piedmont zones of the Northern Caucasus appear to have been largely depopulated until the ninth or eighth century bc[57,70,71], whereas pastoralist groups continued to occupy the high plateaus of the Caucasus Mountains[72].

**Post-Bronze Age adoption of horse milking.** In our dataset, we found no evidence of horse milk consumption until the ninth century bc, when Early Iron Age groups repopulated the North Caucasus steppe and piedmont zones[33,41]. Horses are well adapted to steppe environments, and recent palaeogenomic research has identified the lower Don–Volga region, possibly as early as the mid-sixth millennium bc, as the domestication centre of the DOM2 horses that characterize present-day lineages[43,45]. From the Pleistocene until



Content courtesy of Springer Nature, terms of use apply. Rights reserved

the Bronze Age, horses were hunted on the Pontic–Caspian steppe and have long been symbolically represented in figurines and ritual deposits[28,73]. Horses are also useful for steppe pastoralists because of their digging (*tebenevka*) reflex, which allows them to graze through thick snow deposits, thereby opening up winter pasture for ruminants[48,74,75]. In the North Caucasus, skeletal remains of the ancestors of DOM2 horses are sporadically found in steppe kurgans from the Late Maykop period onwards[43], but the role of horses in these pastoralist societies is unclear. The first undisputed evidence of horse traction dates to ca. 2000 bc at the site of Sintashta east of the Urals, where elaborate horse chariot burials have been found in Middle and Late Bronze Age kurgans[51,76,77]. Earlier Bronze Age wagons, such as those associated with the Late Maykop, Yamnaya and Catacomb cultures, had been pulled by oxen teams[50]. Herding on horseback, which may have begun ca. 2200 bc with the selection of traits suitable for riding[43], would have enabled individual pastoralists to control more livestock at one time and to access pastures across a wider area[75]. Later, horses became particularly prominent in the archaeological record of Early Iron Age Scythians and Sarmatians, who used horses for cavalry[78,79]. In addition to traction and riding, horses can also be exploited for milk, which is traditionally fermented to produce an alcoholic beverage in contemporary Eurasian steppe cultures[80,81]. However, the origin of horse milking is not known. Isotopic evidence from lipids in pottery suggests that Przewalski's horses, reflecting a separate domestication lineage (DOM1)[76], may have been milked as early as the mid-fourth millennium bc at the site of Botai in northern Kazakhstan[76,82]. It is unclear what, if any, influence early milking at Botai had on the management of DOM2 lineages, the ancestors of modern domestic horses. Currently, the earliest proteomic evidence of horse milk consumption comes from two individuals with problematic dates at the Bronze Age site of Kriviyansky IX in the Lower Don region[7] and, later, at the Late Bronze Age site of Uliastai Dood Denzh located in Mongolia, where the dental calculus of an individual dated to ca. 1200 bc with Sintashta-related ancestry yielded evidence of horse milk proteins[2,20]. Despite an apparent early presence of horse milking at Kriviyansky IX, dating to the third, or possibly fourth, millennium bc, we found no other evidence of horse milking in the North Caucasus region during the Early, Middle or Late Bronze Age. Rather, its late appearance in our dataset suggests that horse milking was a highly limited activity while diverse domestication pathways unfolded, and horses were used for various purposes. Horse milking may have been permanently established in the northern Caucasus only after a later reintroduction by pre-Scythian groups during the first millennium bc. Greek texts, such as *The Iliad*, later referred to these pre-Scythian steppe nomads as horse milk drinkers[83].

**Macroregional perspectives on the spread of dairying.** The Pontic–Caspian steppe has long been recognized as a major centre for pastoralist innovation. Here we show that dairying was an early and enduring feature of the pastoralist economy not only in the Northern Caucasus, but also in the South Caucasus. In our dataset, we observed the earliest evidence of milk consumption in the South Caucasus at Alkhantepe, a Late Chalcolithic site with Leilatepe ceramics[84,85]. The contemporaneous Leilatepe and Early Maykop cultures share many features[39,86], but we found that the agropastoralists at Alkhantepe were milking cattle, whereas we observed only sheep milking at Early and Late Maykop sites in the north. Sheep and cattle have different ecological needs, and, in particular, sheep require less water and can survive harsher winters than cattle. As such, environmental factors may have played a role in influencing the selection of dairy livestock in these two regions. During the third millennium bc, it is known that the economic importance of pastoralism increased in the South Caucasus, especially during the Kura–Araxes period[56,87], but we did not have corresponding samples to examine this. Although steppe cultural elements, such as kurgans (burial mounds), had been present in the South Caucasus since the Late Chalcolithic[88], kurgans greatly increased during the Middle Bronze Age[89], and we next observed dairy product consumption at the Middle Bronze Age fortified agropastoral site of Qızqala, with ruminant dairy proteins present in both individuals analysed for this study. Although Middle Bronze Age cultures in both the North and South Caucasus largely became fully mobile to support their herds[90], the inhabitants of Qızqala relied on a more flexible subsistence strategy that included both settlement occupation and seasonal movement of livestock[89,91]. Our results show a reliance on dairy technology for subsistence for these mobile pastoralists. Next, we found evidence of sheep milk consumption by one individual from an intrusive Late Bronze/Early Iron Age burial associated with the Khojaly–Gadabay culture at the Neolithic site of Göytepe. This is the earliest unequivocal evidence of sheep milking in our South Caucasus dataset. Later, during the Greco-Roman era, we observe evidence of sheep, goat and cattle milk at Qabala, a site associated with complex and intensive agriculture as well as with local herding.

Despite cultural interaction with adjacent communities of the Pontic–Caspian steppe, communities in the Oka–Volga–Don forested regions maintained economies based on hunting, gathering and fishing that were particularly suited to local ecozones. Stable isotope studies suggest that this was the prevailing economic strategy until the end of the third millennium bc during the Middle Bronze Age[48,92,93] when Oka–Volga–Don communities transitioned to agropastoralist subsistence[44]. Although populations further to the east, between the Volga River and the Ural Mountains, practiced ruminant dairying from ca. 3000 bc onwards[7], the near-complete lack of evidence for ruminant milk consumption from the seven individuals representing the Oka–Volga–Don region in our study is consistent with a late introduction of ruminant dairying west of the Volga, despite the fact that domesticated animals were introduced in small quantities during the late fourth millennium bc. Here, only one Catacomb-associated individual with cultural links to the steppe zone, recovered from the site of Rovenka, yielded ruminant milk proteins, which were sourced from sheep, goats and cattle.

In parallel to the expansion of pastoralism to the forest–steppe zone, contact and admixture with late farming groups in eastern Europe, such as Cucuteni–Trypillia and Globular Amphora, resulted in a mixed form of agropastoralism with heavy reliance on pastoralism[94], followed by a subsequent eastward expansion of the Corded Ware complex during the third and early second millennia bc, which is also attested by archaeogenetic data[22,95]. This sphere of influence includes Fatyanovo/Balanovo and subsequent Abashevo, Sintashta, Andronovo, and Srubnaya groups[94], and individuals associated with these cultures share very similar genetic profiles. We analysed two individuals linked to the Srubnaya culture at the Middle to Late Bronze Age site of Neplyuyevka in the region east of the Ural Mountains and identified evidence of ruminant milk consumption. Future work combining palaeogenomic and palaeodietary research could help to better clarify the relationships between these populations and the nature and spatio-temporal patterning of dairy technologies in this region.

## Conclusion

Proteomic analysis of human dental calculus has revealed a dynamic trajectory of dairy pastoralism in the North Caucasus steppe and adjacent regions from the Eneolithic to the Greco-Roman periods. Dairying was integral for the spread of animal husbandry by groups crossing the Caucasus mountains from south to north during the Eneolithic, and it was quickly adopted and further developed into an effective and sustainable technology—dairy pastoralism—by neighbouring steppe communities. This innovation forms the basis of the Eurasian steppe lifestyle that continues until today. Initial pastoralist strategies focused on sheep dairying and cattle traction, whereas fully mobile pastoralism arose for the first time during the Yamnaya



Content courtesy of Springer Nature, terms of use apply. Rights reserved

period. Deteriorating climatic conditions challenged steppe herders during the Middle and Late Bronze Ages, who responded by diversifying their set of dairying livestock and expanding their herding range, until the steppe was ultimately abandoned in the mid-second millennium bc. Later, following a centuries-long hiatus, the steppe was repopulated by Early Iron Age pastoralists who practised horse milking. The turbulent third millennium bc, during which vast stretches of Eurasia experienced social and demographic upheaval, is now coming into sharper focus. Climatic pressures and the needs of dairy herds altered how pastoralists used the North Caucasus steppe and may have contributed to the heightened mobility of third-millennium-bc steppe herders, whose descendants spread across Eurasia within the span of only a few centuries. Future research on the genomes of ancient dairying livestock and additional dental calculus proteomes from adjacent steppe populations north of the Black Sea and east of the Urals will help to further clarify the origins and dispersals of dairying breeds and practices that promoted the lasting cultural and subsistence traditions that reshaped the Eurasian steppe zone and profoundly transformed the Bronze Age Eurasian world.

## Methods

**Sampling.** Dental calculus sampling was performed on site at archaeological institutions and museums and in a dedicated ancient biomolecules laboratory at the Max Planck Institute for the Science of Human History (MPI-SHH). Disposable nitrile gloves were worn during collection, and calculus was sampled using dental curettes that were replaced or cleaned with isopropanol between samples. Calculus was collected onto weighing paper and stored in microcentrifuge tubes. Samples were further analysed at the MPI-SHH ancient proteomics laboratory, where they were weighed and subsampled before protein extraction. Approximately 5–13 mg of dental calculus was used for each protein analysis.

**Radiocarbon dating.** A total of 24 new radiocarbon dates were obtained by accelerator mass spectrometry of bone and tooth material at: the Curt-Engelhorn-Zentrum Archäometrie in Mannheim, Germany; the Finnish Museum of Natural History (Hela) in Helsinki, Finland; the Oxford Radiocarbon Accelerator Unit in Oxford, United Kingdom; and the Russian Academy of Sciences in Moscow, Russia. Uncalibrated dates were successfully obtained for all but one tested sample (Supplementary Data 1). An additional 21 previously published radiocarbon dates for individuals in this study were also compiled and analysed, making the total number of directly dated individuals in this study 38 (45 total dates). Dates were calibrated using OxCal v.4.4[96] with the IntCal20 atmospheric curve[97].

**Liquid chromatography–tandem mass spectrometry and data analysis.** Archaeological dental calculus samples from 45 individuals and 5 extraction non-template controls were processed using a filter-aided sample-preparation protocol, modified for ancient proteins (https://doi.org/10.17504/protocols.io.7vwhn7e). In brief, dental calculus was demineralized in 0.5 M EDTA, and proteins were solubilized and reduced using SDS lysis buffer (4% SDS, 0.1 M DTT, 0.1 M Tris HCl). Buffer exchange in 8 M urea and total protein isolation were performed using a Microcon 30 kDa centrifugal filter unit with an Ultracel-30 membrane (Millipore), followed by alkylation using iodoacetamide. Following buffer replacement with triethylammonium bicarbonate (TEAB; 0.05 M), the proteins were digested overnight with sequencing-grade modified trypsin (Promega) at 37 °C. Peptides were recovered by centrifugation in TEAB and acidified with trifluoroacetic acid to pH < 3 and desalted using C18 stage tips (Pierce). Peptides were analysed by liquid chromatography–tandem mass spectrometry using a Q-Exactive mass spectrometer (Thermo Fisher Scientific) coupled to an ACQUITY UPLC M-Class system (Waters AG) at the Functional Genomics Center Zurich of the University/ETH Zurich. Spectra were acquired from 300–1,700 *m/z* with an automatic gain control target of $3 \times 10^6$, a resolution of 70,000 (at 200 *m/z*) and a maximum injection time of 110 ms. The quadrupole isolated precursor ions with a 2.0 *m/z* window, a $5 \times 10^4$ automated gain control value and a maximum fill time of 110 ms. Twelve of the most intense precursor ions for each $MS_1$ scan were fragmented via high collision dissociation with a normalized collision energy of 25, scanned with a resolution of 35,000 (at 200 *m/z*) and a fixed first mass of 200 *m/z*. An intensity threshold of $9.1 \times 10^3$ was applied for $MS_2$ selection, and singly charged ions were excluded. Filter criteria for $MS_2$ selection were an intensity threshold of $9.1 \times 10^3$, and unassigned, singly charged ions were excluded. Selected precursor ions were put onto a dynamic exclusion list for 30 s. For liquid chromatography, the solvent composition at the two channels was 0.1% formic acid in water for channel A and 0.1% formic acid in acetonitrile for channel B. Next, 4 µl of each peptide sample was loaded onto a trap column (Symmetry C18, 100 Å, 5 µm, 180 µm × 20 mm; Waters AG) with a flow rate of 15 µl min$^{-1}$ of 99% solvent A for 60 s at room temperature. Peptides eluting from the trap column were refocused and separated on a C18 column (HSS T3 C18, 100 Å, 1.8 µm, 75 µm × 250 mm; Waters AG). The column temperature was 50 °C. Peptides were separated over 73 min with the following gradient: 8–22% solvent B in 49 min, 22–32% solvent B in 11 min and 32–95% solvent B in 5 min. The column was cleaned with 95% solvent B for 5 min after the separation and re-equilibrated at loading condition for 8 min before initializing the next run. Potential contamination was monitored using extraction blanks.

Tandem mass spectra were converted to Mascot generic files by MSConvert version 3.0.11781 using the 100 most intense peaks in each spectra. All tandem mass spectrometry samples were analysed using Mascot (Matrix Science, version 2.6.0). Mascot was set up to search the SwissProt Release 2019_08 database (560,823 entries) assuming the digestion enzyme trypsin, with automatic decoy option. Mascot was searched with a fragment ion mass tolerance of 0.050 Da and a parent ion tolerance of 10.0 ppm. The number of missed cleavages was specified as one. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Deamidation of asparagine and glutamine and oxidation of methionine and proline were specified in Mascot as variable modifications.

Scaffold version 4.9.0 (Proteome Software Inc.) was used to validate protein and peptide identifications for each sample. Peptide identifications were accepted if they could be established at greater than a 90% probability by the PeptideProphet algorithm. Protein identifications were accepted if they could be established at a greater than 95% probability and contained at least two unique peptides. Probabilities for proteins were assigned using the ProteinProphet algorithm[98]. Proteins that contained similar peptides that could not be differentiated based on tandem mass spectrometry analysis alone were grouped to satisfy the principles of parsimony, and proteins that shared significant peptide evidence were grouped into clusters. Peptide identifications were accepted if they could be established at a greater than 90% probability using the PeptideProphet algorithm[99] with Scaffold delta-mass correction. Individual protein and peptide false discovery rates are listed in Supplementary Data 3.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Raw data files are available through the ProteomeXchange Consortium via the PRIDE partner repository under accession PDX027728. Source data are provided with this paper.

Received: 17 September 2021; Accepted: 7 February 2022;
Published online: 7 April 2022

## References

1. Bleasdale, M. et al. Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nat. Commun.* **12**, 632 (2021).
2. Wilkin, S. et al. Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years. *Nat. Ecol. Evol.* **4**, 346–355 (2020).
3. Evershed, R. P. et al. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* **455**, 528–531 (2008).
4. Salque, M. et al. Earliest evidence for cheese making in the sixth millennium bc in northern Europe. *Nature* **493**, 522–525 (2013).
5. Dunne, J. et al. First dairying in green Saharan Africa in the fifth millennium bc. *Nature* **486**, 390–394 (2012).
6. Hendy, J. et al. Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nat. Commun.* **9**, 4064 (2018).
7. Wilkin, S. et al. Dairying enabled Early Bronze Age Yamnaya steppe expansions. *Nature* **598**, 629–633 (2021).
8. Jeong, C. et al. Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc. Natl Acad. Sci. USA* **115**, E11248–E11255 (2018).
9. Honeychurch, W. & Makarewicz, C. A. The archaeology of pastoral nomadism. *Annu. Rev. Anthropol.* **45**, 341–359 (2016).
10. Weaver, C., Wijesinha-Bettoni, R., McMahon, D. & Spence, L. in *Milk and Dairy Products in Human Nutrition* (eds Muehlhoff, E., Bennett, A. & Macmahon, D.) 103–206 (Food and Agriculture Organization of the United Nations, 2013).
11. Tishkoff, S. A. et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40 (2007).
12. Ségurel, L. & Bon, C. On the evolution of lactase persistence in humans. *Annu. Rev. Genomics Hum. Genet.* **18**, 297–319 (2017).
13. Metheny, K. B. & Beaudry, M. C. *Archaeology of Food: An Encyclopedia* (Rowman & Littlefeld, 2015).
14. Wijesinha-Bettoni, R. & Burlingame, B. in *Milk and Dairy Products in Human Nutrition* (eds Muehlhof, E., Bennett, A. & Macmahon, D.) 41–102 (Food and Agriculture Organization of the United Nations, 2013).
15. Bogaard, A. The archaeology of food surplus. *World Archaeol.* **49**, 1–7 (2017).



Content courtesy of Springer Nature, terms of use apply. Rights reserved

16. Dunne, J., di Lernia, S., Chłodnicki, M., Kherbouche, F. & Evershed, R. P. Timing and pace of dairying inception and animal husbandry practices across Holocene North Africa. *Quat. Int.* **471**, 147–159 (2018).
17. Frachetti, M. D. Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia. *Curr. Anthropol.* **53**, 2–38 (2012).
18. Segurel, L. et al. Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. *PLoS Biol.* **18**, e3000742 (2020).
19. Taylor, W. T. T. et al. Evidence for early dispersal of domestic sheep into Central Asia. *Nat. Hum. Behav.* **5**, 1169–1179 (2021).
20. Jeong, C. et al. A Dynamic 6,000-year genetic history of Eurasia's Eastern Steppe. *Cell* **183**, 890–904.e29 (2020).
21. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
22. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
23. Kovalev, A. A. & Erdenebaatar, D. in *Afanasyevskiy sbornik* [in Russian] (eds Stepanova, N. F. & Polyakov, A. V.) 91–108 (Azbuka, 2010).
24. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
25. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
26. Sjögren, K.-G., Price, T. D. & Kristiansen, K. Diet and mobility in the corded ware of Central Europe. *PLoS ONE* **11**, e0155083 (2016).
27. Papac, L. et al. Dynamic changes in genomic and social structures in third millennium bc central Europe. *Sci. Adv.* **7**, eabi6941 (2021).
28. Benecke, N. Archaeozoological studies on the transition from the Mesolithic to the Neolithic in the North Pontic region. *Anthropozoologica* **25**, 631–641 (1997).
29. Meshveliani, T. et al. Mesolithic hunters at Kotias Klde, Western Georgia: preliminary results. *Paléorient* **33**, 47–58 (2007).
30. Vybornov, A., Kosintesev, P. & Kulkova, M. The origin of farming in the Lower Volga region. *Documenta Praehistorica* **62**, 67–75 (2015).
31. Nishiaki, Y. et al. The Mesolithic-Neolithic interface in the Southern Caucasus: 2016–2017 excavations at Damjili Cave, West Azerbaijan. *Archaeol. Res. Asia* **19**, 100140 (2019).
32. Skourtanioti, E. et al. Genomic history of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* **181**, 1158–1175.e28 (2020).
33. Wang, C.-C. et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
34. Gorelik, A. F., Cybrij, A. & Cybrij, V. Zu kaukasischen und vorderasiatischen Einfüssen bei der Neolithisierung im unteren Donbecken. *Eurasia Antiq.* **20**, 143–170 (2017).
35. Reinhold, A., Belinskiy, A. B. & Korobov, D. S. in *Archaeology of Mountain Landscapes: Interdisciplinary Research Strategies of Agro-Pastoralism in Upland Regions* (ed Garcia-Molsosa, A.) Chapter 9 (State University of New York Press, 2022).
36. Rassamakin, Y. in *Late Prehistoric Exploitation of the Eurasian Steppe* (eds Levine, M. et al.) 59–182 (Univ. Cambridge, 1999).
37. Trifonov, V. in *Adaptation of Paleolithic–Eneolithic Cultures to Changes in the Natural Environment in the North-West Caucasus* [in Russian] (ed Trifonov, V.) 84–93 (Institute for the History of Material Culture RAS, 2009).
38. Kohl, P. L. & Trifonov, V. in *The Cambridge World Prehistory* (eds Renfrew, C. & Bahn, P.) 1571–595 (Cambridge Univ. Press, 2014).
39. Lyonnet, B. in *Les Cultures du Caucase (VIe- IIIe Millénaires avant Notre ère): Leurs Relations avec le Proche-Orient* (ed Lyonnet, B.) 133–161 (CNRS Éditions, 2007).
40. Shishlina, N. in *Counterpoint: Essays in Archaeology and Heritage Studies in Honour of Professor Kristian Kristiansen* (eds Bergerbrant, S. & Sabatini, S.) 53–60 (BAR International, 2013).
41. Knipper, C. et al. Diet and subsistence in Bronze Age pastoral communities from the southern Russian steppes and the North Caucasus. *PLoS ONE* **15**, e0239861 (2020).
42. Benecke, N. in *The Oxford Handbook of Zooarchaeology* (eds Albarella, U. et al.) 319–332 (Oxford Univ. Press, 2017).
43. Librado, P. et al. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* **598**, 634–640 (2021).
44. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. (Princeton Univ. Press, 2010).
45. Guimaraes, S. et al. Ancient DNA shows domestic horses were introduced in the southern Caucasus and Anatolia during the Bronze Age. *Sci. Adv.* **6**, eabb0030 (2020).
46. Staubwasser, M. & Weiss, H. Holocene climate and cultural evolution in late prehistoric–early historic West Asia. *Quat. Res.* **66**, 372–387 (2006).
47. Shishlina, N. *Reconstruction of the Bronze Age of the Caspian Steppes: Life Styles and Life Ways of Pastoral Nomads* (Archaeopress, 2008).
48. Shishlina, N., Sevastyanov, V. & Kuznetsova, O. Seasonal practices of prehistoric pastoralists from the south of the Russian plain based on the isotope data of modern and archaeological animal bones and plants. *J. Archaeol. Sci. Rep.* **21**, 1247–1258 (2017).
49. Hambleton, E. & Maltby, M. *The Animal Bones from Excavations in Meshoko Cave in the Northern Caucasus* (Bournemouth Univ. & St Petersburg State Univ., 2016); http://eprints.bournemouth.ac.uk/24860/1/MeshokoCaveReportJuly2016.pdf
50. Reinhold, S. et al. in *Appropriating Innovations: Entangled Knowledge in Eurasia, 5000–1500* BC (eds Stockhammer, P. & Maran, J.) 78–97 (Oxbow Books, 2017).
51. Chechushkov, I. V. & Epimakhov, A. V. Eurasian steppe chariots and social complexity during the Bronze Age. *J. World Prehistory* **31**, 435–483 (2018).
52. Shishlina, N., Orfinskaya, O. V. & Hommel, P. Bronze Age wool textile of the Northern Eurasia: new radiocarbon data. *Nanotechnol. Russ.* **15**, 629–638 (2020).
53. Ethier, J. et al. Earliest expansion of animal husbandry beyond the Mediterranean zone in the sixth millennium bc. *Sci. Rep.* **7**, 7146 (2017).
54. Stojanovski, D. et al. Living of the land: terrestrial-based diet and dairying in the farming communities of the Neolithic Balkans. *PLoS ONE* **15**, e0237608 (2020).
55. McClure, S. B. et al. Fatty acid specific $\delta^{13}C$ values reveal earliest Mediterranean cheese production 7,200 years ago. *PLoS ONE* **13**, e0202807 (2018).
56. Herrscher, E. et al. Dietary inferences through stable isotope analysis at the Neolithic and Bronze Age in the southern Caucasus (sixth to first millenium bc, Azerbaijan): from environmental adaptation to social impacts. *Am. J. Phys. Anthropol.* **167**, 856–875 (2018).
57. Reinhold, S., Korobov, D. & Belinskij, A. B. *Landschaftsarchäologie im Nordkaukasus. Studien zu einer Neu Entdeckten bronzezeitlichen Kulturlandschaft im Hochgebirge des Nordkaukasus* Vol. 38 (Habelt-Verlag, 2017).
58. Chazin, H. *The Politics of Pasture: The Organization of Pastoral Practices and Political Authority in the Late Bronze Age in the South Caucasus* (Univ. Chicago, 2016).
59. Uerpmann, M. & Uerpmann, H. P. Bronze and Iron Age animal economy at Didi-gora and Tqisbolo-gora (Kakhetia, Georgia). *Archäol. Mittl. aus Iran und Turan* **40**, 169–264 (2008).
60. Warinner, C. et al. Direct evidence of milk consumption from ancient human dental calculus. *Sci. Rep.* **4**, 7104 (2014).
61. Ricci, A., D'Anna, M. B., Lawrence, D., Helwing, B. & Aliyev, T. Human mobility and early sedentism: the Late Neolithic landscape of southern Azerbaijan. *Antiquity* **92**, 1445–1461 (2018).
62. Berthon, R. Past, current and future contribution of zooarchaeology to the knowledge of the Neolithic and Chalcolithic cultures in South Caucasus. *Stud. Cauc. Archaeol.* **2**, 4–30 (2014).
63. Piotrovskij, J. J. in *Der Kaukasus zwischen Osteuropa und Vorderem Orient in der Bronze und Eisenzeit* (eds Kašuba, M. T. et al.) Vol. 19 (Dietrich Reimer Verlag, 2020).
64. Korenevsky, S. N. *The Most Ancient Farmers and Stockbreeders of the Pre-Caucasus Region: The Maikop–Novosvobodnoe Community, Problems of Internal Typology and Symbolism of Funerary Practices* [in Russian] (Nauka, 2004).
65. Merpert, N. Y. A. The most ancient stockbreeders of the Volga-Ural interfluve [in Russian] (Nauke, 1974).
66. Kaiser, E. *Das Dritte Jahrtausend im Osteuropäischen Steppenraum: Kulturhistorische Studien zu Prähistorischer Subsistenzwirtschaft und Interaktion mit Benachbarten Räumen*. Vol. 37 (Freie Univ. Berlin, 2019).
67. Connor, S. E. & Kvavadze, E. V. Modelling late Quaternary changes in plant distribution, vegetation and climate using pollen data from Georgia, Caucasus. *J. Biogeogr.* **36**, 529–545 (2009).
68. Davis, B. A. S., Brewer, S., Stevenson, A. C. & Guiot, J. The temperature of Europe during the Holocene reconstructed from pollen data. *Quat. Sci. Rev.* **22**, 1701–1716 (2003).
69. Borisov, A. V., Demkina, T. S. & Demkin, V. A. *Paleoenvironments and the Climate of Yerengei in the Bronze Age, IV-II Millennium* bc [in Russian] (Nauka, 2006).
70. Malashev, V. Y. in *Traditions of the Middle Sarmatian Culture in Burial Mounds of the Caspian Dagestan in the Second Half of the Second–First Half of the Fifth Century* ad [in Russian] Vol. 35 85–97 (Bulletin of Volga State Univ., 2015).
71. Erlikh, V. R. in *Achaemenid Impact in the Black Sea: Communication of Powers* (eds Nieling, J. & Rehm, E.) 47–66 (Aarhus Univ. Press, 2010).
72. Reinhold, S., Belinskij, A. B. & Atabiev, B. H. in *Der Kaukasus zwischen Osteuropa und Vorderem Orient in der Bronze und Eisenzeit* (eds Kašuba, M. T. et al.) Vol. 19 (Dietrich Reimer Verlag, 2020).
73. Anthony, D. W. & Brown, D. R. Eneolithic horse exploitation in the Eurasian steppes: diet, ritual and riding. *Antiquity* **74**, 75–86 (2000).
74. Taylor, W. T. T. et al. Early pastoral economies and herding transitions in Eastern Eurasia. *Sci. Rep.* **10**, 1001 (2020).



Content courtesy of Springer Nature, terms of use apply. Rights reserved

# 7. Manuscript D

**Manuscript Nr. :** (Manuscript D)
**Title of the Manuscript:** Exotic foods reveal contact between South Asia and the Near East during the 2nd millennium BCE
**Authors:** Ashley Scott, Robert C. Power, Victoria Altmann-Wendling, Michal Artzy, Mario A. S. Martin, Stefanie Eisenmann, Richard Hagan, Domingo C. Salazar-García, Yossi Salmon, Dmitry Yegorov, Ianir Milevski, Israel Finkelstein, Philipp W. Stockhammer, Christina Warinner
**Bibliographic Information:** Scott A, Power RC, Altmann-Wendling V, Artzy M, Martin M. et al. (2021) Exotic foods reveal contact between South Asia and the Near East during the second millennium BCE. *Proceedings of the National Academy of Sciences* USA 118(2) e2014956117
**The candidate is:** ☒ First author, ☐ Co-first author, ☐ Corresponding author, ☐ coauthor
**Status:** Published
**Proportion (in %) of authors in the publication:** (to be indicated from 20%)

| Author | Concept | Data analysis | Experimental | Manuscript composition | Material provision |
|---|---|---|---|---|---|
| Scott, A | 25% | 50% | 50% | 50% | |
| Power, RC | 25% | 50% | 50% | | |
| Altmann-Wendling, V | | | | | |
| Artzy, M | | | | | |
| Martin, M | | | | | |
| Eisenmann, S | | | | | |
| Hagan, R | | | | | |
| Salazar- Garcia, D | | | | | |
| Salmon, Y | | | | | 30% |
| Yegorov. D | | | | | |
| Milevski, I | | | | | |
| Finkelstein, I | | | | | |
| Stockhammer, P | 25% | | | 20% | 40% |
| Warinner C | 25% | | | 20% | |

# Exotic foods reveal contact between South Asia and the Near East during the second millennium BCE

Ashley Scott[a,b], Robert C. Power[b,c], Victoria Altmann-Wendling[b,d], Michal Artzy[e], Mario A. S. Martin[f], Stefanie Eisenmann[a,b], Richard Hagan[a,b], Domingo C. Salazar-García[g,h], Yossi Salmon[e], Dmitry Yegorov[i], Ianir Milevski[i], Israel Finkelstein[f], Philipp W. Stockhammer[a,b,1], and Christina Warinner[a,j,1]

[a]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07743 Jena, Germany; [b]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig Maximilian University Munich, 80539 Munich, Germany; [c]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany; [d]Institute for Egyptology, Julius-Maximilians-University Würzburg, 97070 Würzburg, Germany; [e]Leon Recanati Institute for Maritime Studies, University of Haifa, 3498838 Haifa, Israel; [f]Institute of Archaeology, Tel Aviv University, 69978 Tel Aviv, Israel; [g]Ikerbasque-Basque Foundation for Science, Grupo de Investigación en Prehistoria IT-1223-19 (Universidad del País Vasco/Euskal Herriko Unibertsitatea), 01006 Vitoria-Gasteiz, Spain; [h]Departament de Prehistòria, Arqueologia i Història Antiga, University of València, 46010 València, Spain; [i]Archaeological Research Department, Archaeological Division, Israel Antiquities Authority, Jerusalem 91004, Israel; and [j]Department of Anthropology, Harvard University, Cambridge, MA 02138

Edited by Dolores R. Piperno, Smithsonian Institution, Washington, DC, and approved November 10, 2020 (received for review July 21, 2020)

**Although the key role of long-distance trade in the transformation of cuisines worldwide has been well-documented since at least the Roman era, the prehistory of the Eurasian food trade is less visible. In order to shed light on the transformation of Eastern Mediterranean cuisines during the Bronze Age and Early Iron Age, we analyzed microremains and proteins preserved in the dental calculus of individuals who lived during the second millennium BCE in the Southern Levant. Our results provide clear evidence for the consumption of expected staple foods, such as cereals (Triticeae), sesame (*Sesamum*), and dates (*Phoenix*). We additionally report evidence for the consumption of soybean (*Glycine*), probable banana (*Musa*), and turmeric (*Curcuma*), which pushes back the earliest evidence of these foods in the Mediterranean by centuries (turmeric) or even millennia (soybean). We find that, from the early second millennium onwards, at least some people in the Eastern Mediterranean had access to food from distant locations, including South Asia, and such goods were likely consumed as oils, dried fruits, and spices. These insights force us to rethink the complexity and intensity of Indo-Mediterranean trade during the Bronze Age as well as the degree of globalization in early Eastern Mediterranean cuisine.**

proteomics | Bronze Age | Eastern Mediterranean | spice trade | early globalization

Long-distance trade across Eurasia has played a major role in connecting distant societies throughout recorded history, with the silk and spice trade being emblematic for early globalization (1–3). Recent decades of archaeological research have demonstrated the deep prehistory of these trans-Eurasian exchange networks and support a much earlier onset of globalization, known as "Bronzization" (4), which traces its roots to the Bronze Age during the third millennium BCE. The increasing importance of bronze served as a major impetus for the establishment of extensive trade contacts (5), which were largely driven by the uneven distribution of highly valued raw materials, such as tin, carnelian, and lapis lazuli (6–8). In addition to these raw materials, finished objects—as well as technologies, practices, and knowledge—were also conveyed over unprecedented distances. The major corridors of exchange connected Eurasia both by land across the Eurasian steppes and Iranian Plateau and by sea from India to the Near East via both the Persian Gulf and the Red Sea (9, 10). These networks served to link the major Bronze Age river valley societies of Egypt, Mesopotamia, the Indus Valley, and Central China to neighboring cultures in the Levant, the Arabian Peninsula, the Iranian Plateau, and the Central Asian oases (11–13). Such trade networks also further extended into Anatolia, the Aegean, and throughout South and East Asia. Despite periodic disruptions of some trade routes (14), the intensity of exchange gained momentum during the Middle and Late Bronze Ages of the second millennium BCE (15–17). During this period, bronze was produced on a large scale across Eurasia (4, 5, 18), and urban societies and early states linked by these trade routes developed a rapidly growing interest in exotic goods, including plant and animal products. In the early first millennium BCE, such trade networks had effectively linked West and East Asia, and several economically important crops had become widely dispersed throughout the continent (Fig. 1, and *SI Appendix*, Fig. S1).

**Evidence for Long-Distance Trade of Material Goods and Animals.** During the second millennium BCE, textual evidence in the Near East attests to a large amount of goods being transported over great distances. For example, cuneiform tablets from the Assyrian trade post of Kaneš (Kanesh) in Anatolia record caravans of hundreds of donkeys regularly transporting goods between the Mesopotamian city of Aššur (Assur) and Central Anatolia during the 19th and18th centuries BCE (19, 20). Starting with the expansion policy of Thutmosis III (15th century BCE) into the Levant, the flow of goods and people in West Asia

ANTHROPOLOGY

### Significance

Here we report the identification of staple and exotic food remains in Bronze and Early Iron Age dental calculus from the Southern Levant. The analysis of dietary plant microremains and proteins sheds new light on consumed exotic foods from South and East Asia during the second millennium BCE. We provide the earliest direct evidence in the Mediterranean to date for the consumption of sesame, soybean, probable banana, and turmeric. The recovery and identification of diverse foodstuffs using molecular and microscopic techniques enables a new understanding of the complexity of early trade routes and nascent globalization in the ancient Near East and raises questions about the long-term maintenance and continuity of this trade system into later periods.

Author contributions: A.S., R.C.P., M.A., Y.S., P.W.S., and C.W. designed research; A.S., R.C.P., S.E., R.H., and D.C.S.-G. performed research; A.S., R.C.P., V.A.-W., M.A.S.M., D.Y., I.M., I.F., P.W.S., and C.W. analyzed data; and A.S., R.C.P., V.A.-W., P.W.S., and C.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence may be addressed. Email: philipp.stockhammer@lmu.de or warinner@fas.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2014956117/-/DCSupplemental.

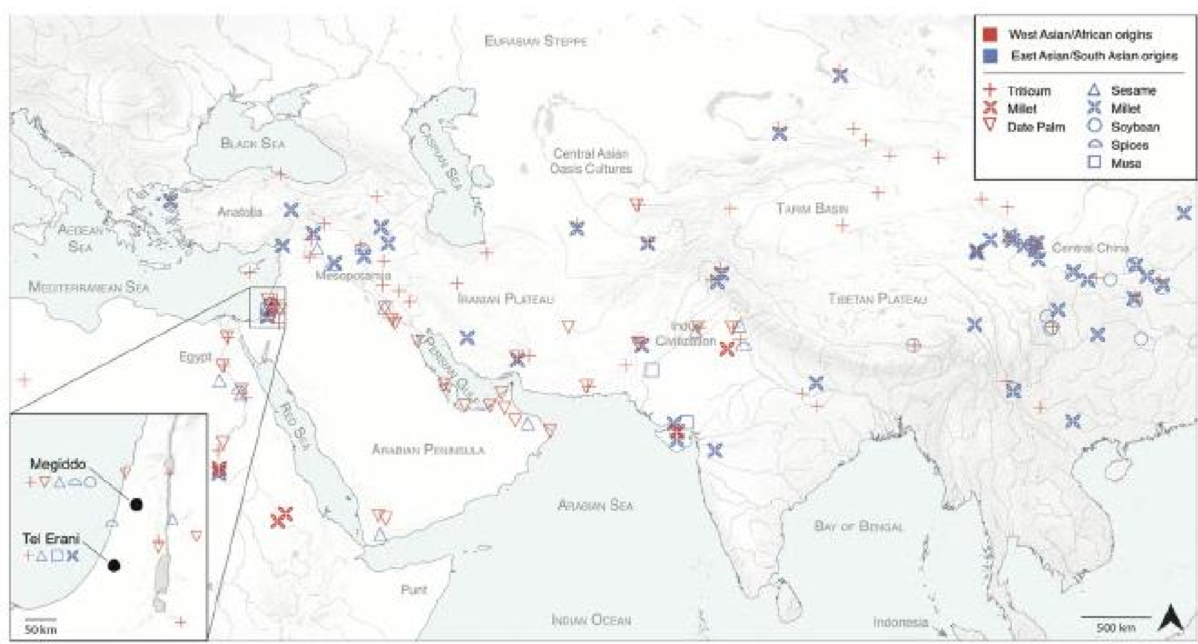



**Fig. 1.** Map of representative archaeobotanical evidence for the spread and trade of food crops prior to 500 BCE. See *SI Appendix* for data sources. Map *Inset* shows the location of the sites of Megiddo and Tel Erani on the southern Levantine coast; new dietary finds reported in this study are indicated for each site.

intensified, which is documented most prominently in the famous Amarna letters (21). Ancient literary sources and illustrations also reference long-distance journeys to foreign lands to obtain exotic goods, such as ivory, ostrich eggshells, ebony, and frankincense. Among the most well-known of these accounts is an expedition initiated by the Egyptian queen Hatshepsut to the land of Punt (probably located in the Horn of Africa region) in the 15th century BCE (22, 23). In addition, evidence for long-distance trade between the Near East and the Indian subcontinent is growing for the second and third millennium BCE (8, 24), and includes written accounts (25–27), seals and stone weights (8, 28–30), shell, lapis lazuli, and carnelian jewelry (8, 24, 31, 32), and timber and ivory (33, 34).

Live animals were also transported long distances. Zooarchaeological and isotopic analyses have identified the movement of donkeys from Egypt to the Southern Levant during the third millennium BCE (35), and ancient DNA analysis has documented the transport of pigs from Italy to the southern Greek mainland (36) and from the Aegean to the Southern Levant during the second millennium BCE (37). In addition, depictions of zebu (*Bos taurus indicus*), a humped subspecies of cattle native to South Asia (38), have been found in Mesopotamia as early as the third millennium BCE (39), and a clear depiction of a zebu (Fig. 2) appears on a bichrome ceramic vessel from the southern Levant dating to 16th century BCE, the same period when depictions of zebu also become common in Egypt (39). Zooarchaeological study of faunal remains and genetic data further confirm the presence of zebu cattle in the Levant during the second millennium BCE (37, 40), and genetic evidence suggests possible taurine–zebu hybridization at the site of Megiddo at approximately 900 BCE (37). Monkeys depicted in 18th and 17th century BCE frescoes at Akrotiri on the Aegean island of Thera were recently identified as South Asian gray langurs (*Semnopithecus* sp.), with a probable origin in the Indus valley (41), and the identification of dermestid khapra beetles (*Trogoderma granarium*) that had infested a wheat deposit from a Middle Kingdom (2050 to 1710 BCE) Egyptian tomb at the site of el-Gebelein points in the same direction (42). At the same time, other animals, such as domestic chicken (*Gallus gallus domesticus*) were brought from East Asia to the Near East (possibly via South Asia) and displayed as exotic curiosities and used for cockfighting. Sporadic evidence of chicken occurs in Anatolia, Iran, Syria, and Egypt as early as the third millennium BCE, and by the late first millennium BCE some sites in the Levant appear to have economically specialized in chicken husbandry (43).

**Evidence for Long-Distance Trade of Culinary and Economic Plants.** Although evidence for the movement of durable goods and animals is richly attested by historical and archaeological evidence, direct evidence for culinary and economic plants is more limited. Plant remains are highly perishable, and only certain conditions lead to macrobotanical preservation, which may be biased with respect to plant type or part (44). Nevertheless, plant remains, such as charred seeds, have documented the eastward spread of wheat and barley across Eurasia during and after the late fourth millennium BCE (45), as well as the westward spread of millet cultivation from East Asia during the third millennium BCE (9, 46, 47). Macrobotanical and microbotanical evidence also confirm the establishment of citron (*Citrus medica*), a fruit tree of South/Southeast Asian origin (48, 49), as an important crop in the Levant and Egypt by the first millennium BCE (50–52), although it may have been first introduced into the Eastern Mediterranean as early as the fourth millennium BCE (53, 54). Melon (*Cucumis melo*), another important crop of South Asian origin (55), was also cultivated in the Near East during the Bronze Age. Textual references to melons appear in third millennium BCE Sumerian texts, and melons and cucurbits are depicted in Egyptian tombs from the Old Kingdom onwards (54, 56, 57). Cucurbit seeds have been reported in Near Eastern archaeological contexts as early as the sixth century BCE (54), but finds prior to the first millennium BCE are less secure in their taxonomic assignment (56).

Fig. 2. Iconographic representation of a zebu on a bichrome ceramic vessel from Tel Gerisa, Israel, 16th century BCE (photo courtesy of Zeev Herzog).



Beyond grains and fruits, there is also growing evidence for a spice trade between South Asia and the Eastern Mediterranean. This is supported by recent findings from organic residue analysis, including evidence for vanillin (58) and possibly also cinnamon (59), nutmeg, and jasmine (60), although evidence for the latter three spices requires further confirmation. Whereas many aspects of this early trade remain unknown, some extraordinary finds leave no doubt that an Indo-Mediterranean spice trade already existed during the Bronze Age (6, 9, 59). Peppercorns used in the mummification of Ramses II in 1213 BCE, for example, are native to southern India (61–63), and cloves, originally from Indonesia, were found at Terqa, Syria dating to 1720 BCE (64, 65), having likely followed an indirect route to Mesopotamia via South Asian trade routes (30, 66). Both of these examples highlight the extent of Indian Ocean trade during this period, despite the declining influence of the Indus Valley and the restructuring of political networks throughout the region (33).

**Emerging Picture of a Dynamic and Complex Exchange Network.** While the details of Bronze Age trade remain patchy, the overall evidence points toward the existence of a dynamic and complex exchange network connecting the Mediterranean with South Asia and beyond during the Middle Bronze Age (approximately 2000 to 1550 BCE), Late Bronze Age (approximately 1550 to 1200/1150 BCE), and Early Iron Age/Iron Age I (approximately 1200/1150 to 1000 BCE). Here we aim to explore the transformation of eastern Mediterranean cuisine as a consequence of Bronze Age globalization by analyzing microscopic and molecular traces of food remains in human dental calculus, a calcified form of dental plaque, from the Southern Levant during the second millennium BCE (Table 1). We focus on two sites, the Middle to Late Bronze Age urban center of Megiddo (17th to 15th centuries BCE) and the Early Iron Age site of Tel Erani (11th century BCE) (Fig. 1). In total, we analyze dental calculus from 16 individuals: 13 from Megiddo (*SI Appendix*, Fig. S2) and 3 from Tel Erani (*SI Appendix*, Fig. S3).

During the Middle and Late Bronze Age, Megiddo was a major urban center in the Southern Levant, and it was embedded within long-distance networks and ruled by local kings. We selected Megiddo because it had already yielded suggestive evidence for exchange with South Asia in the form of both zebu genetic evidence (37) and vanillin residues (58). In this study, we analyze individuals from a variety of mortuary contexts, including pit burials, brick-lined pit burials, pithos burials, masonry-constructed collective tombs, and a recently excavated Middle Bronze Age royal tomb containing the so-called "king" (MGD001) and "queen" (MGD002) of Megiddo (*SI Appendix*).

Dating to ∼500 y later in time, the Tel Erani cemetery is one of the few Early Iron Age cemeteries excavated to date in Israel (*SI Appendix*). Although the related settlement is less understood, the cemetery has been associated with the "Philistine" occupation at the Southern Levant from the 12th century BCE onwards. In contrast to Megiddo, where several individuals sampled by us derived from high- or highest-status burial contexts, the individuals from Tel Erani instead appear to represent the rural general population of the Early Iron Age in this region.

By analyzing the two sites, we aim to gain a broader perspective on Levantine cuisines during the second millennium BCE, a period that witnessed the blossoming of Middle Bronze Age city states, periods of Egyptian domination and retreat during the Late Bronze Age, and the emergence of the so-called Philistines at the onset of the Early Iron Age during the 12th century BCE

## Results

Plant remains were abundant in the Megiddo and Tel Erani dental calculus, and plant microremains and proteins were observed in all 16 dental calculus samples. Of these, probable dietary microremains were identified in all 16 analyzed samples (*SI Appendix*, Table S1), and included phytoliths consistent with wheat (*Triticum*), panicoid/millet (Panicoideae), and date palm (*Phoenix* sp.). Dietary proteins were observed in 5 of 14 analyzed specimens (*SI Appendix*, Table S2), and consisted of 19 dietary

**Table 1. Overview of dietary findings**

| Individual | Burial | Context* | Microremains | Dietary proteins |
|---|---|---|---|---|
| Megiddo, Middle Bronze Age III to Late Bronze Age I, $n = 13$ | | | | |
| MGD001 | Tomb 50 | "King"; elite masonry chamber tomb (triple burial); ca. 1650–1550 BCE | Poaceae, cf. *Triticum*, *Phoenix* (date palm), Arecaceae unspecific (palm), bark, eudicot | None |
| MGD002 | Tomb 50 | "Queen"; elite masonry chamber tomb (triple burial); ca. 1650–1550 BCE | Poaceae, eudicot | None |
| MGD006 | Tomb 16/H/45 | Double pit burial; ca. 1550–1450 BCE | Poaceae, eudicot | None |
| MGD007 | Tomb 16/H/45 | Double pit burial; ca. 1550–1450 BCE | Poaceae | None |
| MGD008 | Tomb 12/K/89 | Pit burial; 1496–1320 BCE | Poaceae, eudicot | None |
| MGD009 | Tomb 12/K/96 | Double pit burial; ca. 1650–1400 BCE | Poaceae, cf. *Triticum*, Triticeae, eudicot | None |
| MGD010 | Tomb 12/K/96 | Double pit burial; 1638–1413 BCE | Poaceae, Triticeae, eudicot | None |
| MGD011 | Tomb 14/K/ 119, lower | Stone-lined cist, exotic grave goods; 1688–1535 BCE | Poaceae, cf. *Triticum*, *Phoenix* (date palm), eudicot | *Sesamum*, *Triticum/Aegilops* |
| MGD013 | Tomb 10/K/118 | Triple pithos burial; ca. 1650–1400 BCE | Poaceae, *Phoenix* (date palm), eudicot, bark | Not analyzed |
| MGD016 | Tomb 14/K/49 | Double brick-lined pit Burial; 1509–1432 BCE | Poaceae, eudicot | None |
| MGD017 | Tomb 100 | Masonry chamber tomb with corbelled roof, many commingled individuals; exotic grave goods; ca. 1650–1400 BCE | Poaceae, Cyperaceae, *Phoenix* (date palm), Arecaceae (palm), eudicot | None |
| MGD018 | Tomb 100 | Masonry chamber tomb with corbelled roof, many commingled individuals; exotic grave goods; ca. 1630–1550 BCE | Poaceae, eudicot, Triticeae | *Curcuma*, *Glycine* |
| MGD021 | Tomb 50 | Elite masonry chamber tomb (triple burial); ca. 1650–1550 BCE | Poaceae, Triticeae, *Phoenix* (date palm), eudicot | Not analyzed |
| Tel Erani, Early Iron Age, $n = 3$ | | | | |
| ERA005 | Burial L2091 | Burial offering of one juglet; ca. 1100–1000 BCE | Poaceae | *Sesamum* |
| ERA017 | Burial L2160 | Burial offering of one Flask; ca. 1100–1000 BCE | Poaceae, Cyperaceae, eudicot, panicoid cf. millet | *Musa* |
| ERA023 | Burial L2181 | No burial offerings; ca. 1100–1000 BCE | Poaceae, eudicot | *Sesamum* |

*Dates are provided as relative dates (marked with "ca.") or calibrated radiocarbon dates (2σ). See Dataset S1.

proteins from cereals (*Triticum*/*Aegilops*), oilseeds (*Sesamum*, *Glycine*), fruits (*Musa*), and spices (*Curcuma*).

**Microremains.** Microremains were analyzed from all 16 individuals, including 2 individuals (MGD013 and MGD021) with insufficient calculus to conduct proteomic analysis (*SI Appendix*, Table S1 and Datasets S2 and S3). Representative examples of the dietary microremains observed in the Megiddo and Tel Erani dental calculus are provided in Fig. 3. Both the Megiddo and Tel Erani individuals produced microremains assemblages consisting of three main components: 1) Dietary microremains, most notably phytoliths (84 total morphotypes, 4,983 phytoliths), but also starches (333 granules); 2) nondietary remains, such as fibers; and 3) ambiguous microremains, such as fungal particles, charcoal particles, and other remains of unknown or ambiguous origin (Dataset S2). The abundance of microremains observed across individuals was highly variable (*SI Appendix*, Fig. S4), ranging from 9 (MGD007) to 1,795 (MGD001).

Phytoliths, a robust and abundant type of plant microfossil, allow the identification of vascular plants, and among angiosperms (flowering seed plants), they primarily form in monocots, particularly in members of the plant family Poaceae (grasses, including cereals). In contrast, magnoliids, eudicots, and other angiosperms (which include most fruits and vegetables) tend to form few phytoliths, while gymnosperms (nonflowering seed plants) produce even fewer (67). The dental calculus phytoliths observed in this study corresponded to these expected proportions (*SI Appendix*, Figs. S4 and S5) and were predominantly derived from the leaves, stems, and husks of wild or domestic Poaceae (*SI Appendix*, Fig. S6), totaling 2,551 phytoliths, plus an additional 916 unspecific Poaceae phytoliths (*SI Appendix*, Table S1). Smaller numbers of morphotypes from other monocots such as Cyperaceae (sedges) and Arecaceae (palms), as well as eudicot fruit/leaf phytoliths and nondiagnostic bark (angiosperm or gymnosperm), were identified in 13 individuals (*SI Appendix*, Fig. S7), totaling 245 phytoliths (*SI Appendix*, Table S1). Starch granules were rare in all samples, except ERA23 (*SI Appendix*, Fig. S4), even in samples not exposed to heat during protein extraction, and few could be identified. However, individual starch granules consistent with Triticeae were found in ERA023, MGD018, and MGD021 (*SI Appendix*, Table S1). In addition, we found 26 other types of microremains whose origins are either ambiguous (pollen, fungal particles, diatoms, foraminifera, sponge spicules, charcoal, fibers, insect fragments) or possible contaminants (skin scales) (Datasets S2 and S3). These microremains were not further analyzed.

Among phytoliths assigned to Poaceae, all morphotypes that form in grasses were represented, including those forming in epidermal short cells, long cells, bulliforms, hairs, papilla, and stoma (Dataset S2), in both single-cell and articulated forms. Grass phytoliths predominantly belonged to the pooid group ($n = 432$), but a smaller number of single-cell short cells could be assigned to the chloridoid ($n = 51$) and panicoid ($n = 10$) grass types (*SI Appendix*, Fig. S8). Although most phytoliths could not be assigned to specific wild or domesticated sources, the assemblage is consistent with grain consumption, and 105 articulated dendritic types were identified in 8 individuals (ERA017, MGD001, MGD002, MGD009, MGD010 and MGD011, MGD017, and MGD021). Among these, 10 articulated husk phytoliths consistent with *Triticum* (wheat) were identified in MGD001, MGD009, and MGD011, and 5 short cells (wide lobed bilobates) deriving from Panicoideae grasses (e.g., millets) were found in ERA017 (*SI Appendix*, Table S1 and Dataset S2).

Cones representing sedge (Cyperaceae) leaf were found in MGD001, MGD017, MGD021, and ERA017 (*SI Appendix*, Fig. S7 and Table S1). Such cones form in a variety of sedge cells, including sedge achenes and achene bracts cells. Palm (Arecaceae) phytoliths were identified in five individuals (*SI Appendix*, Table S1).

A total of 12 globular rugulate/echinate phytoliths of date palm (*Phoenix*) were observed in MGD001, MGD011, MGD013, MGD017, and MGD021, and phytoliths from another unknown species of palm were also detected in MGD001 and MGD017 (Datasets S1 and S2). Globular echinate palm leaf phytoliths are known to be a particularly resilient morphotype (68). Finally, 217 plate, jigsaw, sclereid, tracheid, and related phytoliths from eudicot epidermal tissue were identified in 14 individuals, including a single Megiddo individual (MGD001) with 3 jigsaw morphotypes displaying protuberance decorations thought to originate from the eudicot epidermal tissue of fruit and seeds (*SI Appendix*, Fig. S7 and Datasets S2 and S3) (69).

**Proteomics.** Total protein was extracted from the dental calculus of 14 individuals. Protein recovery was variable, but all samples yielded proteins typical of an oral microbiome (*SI Appendix*, Fig. S9 and Dataset S4) and contained damage-associated modifications (N,Q deamidation) (*SI Appendix*, Table S3) consistent with ancient samples. Dietary proteins were identified in 5 individuals and consisted of 19 proteins from 5 plants of known dietary importance: wheat (*Triticum*/*Aegilops*), sesame (*Sesamum*), soybean (*Glycine*), banana (probable *Musa*), and turmeric (*Curcuma*) (Fig. 4 and *SI Appendix*, Fig. S10 and Table S2). Detailed information regarding the identified peptides and peptide spectral matches (PSMs) for each protein is provided in Dataset S5.

Wheat proteins were identified in a single individual from Megiddo (MGD011) and consisted of two major seed proteins: α-amylase inhibitor (AAI) and low molecular weight (LMW) glutenin. AAI makes up ~4% of the protein content of wheat and is highly resistant to heat or proteolytic digestion (70). The identification of AAI was supported by three PSMs, with two PSMs matching specifically to either *Triticum aestivum* or *Aegilops tauschii* (a wild progenitor of *T. aestivum*), while the third peptide was less specific and is also found in *Hordeum* (barley). LMW glutenin, a major gluten protein, was identified with the support of 3 peptides and 13 PSMs. Gluten proteins make up ~80% of the total proteins in whole wheat flour, and LMW glutenin accounts for 20 to 35% of the gluten protein content (71).

Sesame proteins were identified in one individual from Megiddo (MGD011) and two from Tel Erani (ERA005, ERA023) and consisted of 2 proteins supported by a total of 29 peptides and 78 PSMs: 11S globulin and 2S albumin seed storage proteins. These two seed storage proteins are tissue-specific and expressed during seed maturation. Together, they make up most of the protein in sesame seeds, accounting for 60 to 70% and 15 to 25% of total seed protein, respectively (72).

Soybean was identified in a single Megiddo individual (MGD018) and is supported by 30 PSMs specific to the proteins glycinin, an 11S storage protein, and β-conglycinin, a 7S storage protein. These two proteins are the primary storage proteins in soybean seeds, and together they make up ~65% of the total soybean protein content (73). In addition to these two seed storage proteins, we also identified two peptides supported by two PSMs specific to soybean sucrose-binding protein, another member of the seed storage protein superfamily.

We identified endo-1,3-β-glucanase, an enzyme important in fruit ripening, in a single individual from Tel Erani (ERA017). This protein was supported by two peptides, of which one is highly specific to banana (*Musa* or *Ensete*), while the second is found in *Musa* and a number of other flowering plants, but not *Ensete*. Of the two plants, only members of *Musa* produce edible fruits, while *Ensete* (Abyssinian banana) is consumed for its starchy pseudostem and corm. The fruit of ripe bananas contains few proteins but relatively high concentrations of β-glucanases (74), which are highly resistant to heat and proteolytic degradation (75). Because this protein is expressed in fruit and peel and increases in abundance as the fruit ripens (76, 77), we tentatively identify this enzyme as originating from *Musa*. However, the enzyme is also expressed in generalized stress response, and thus is also present in diseased and injured plant tissues. Taken together, the identification of *Musa* is most strongly supported, but *Ensete* cannot be fully excluded.



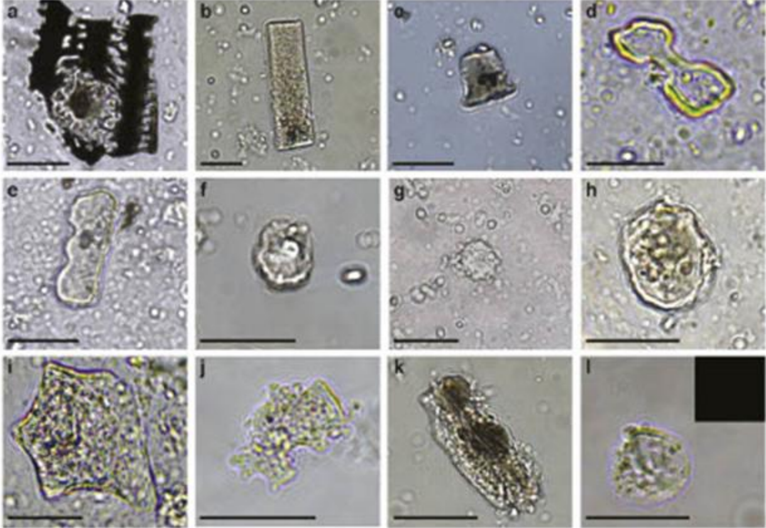



**Fig. 3.** Microremains in Megiddo and Tel Erani dental calculus. (*A*) Articulated Poaceae husk phytolith, identified as wheat (MGD001). (*B*) Poaceae stem/leaf phytolith (MGD001). (*C*) Poaceae short cell rondel (MGD001). (*D*) Wide-lobed bilobate short cell identified as panicoid (ERA017). (*E*) Poaceae polylobate short cell (MGD001). (*F*) Cone phytolith identified as sedge leaf (MGD018). (*G*) Spheroid echinate identified as date palm (MGD001). (*H*) Spheroid echinate phytolith, identified as nondiagnostic palm (MGD001). (*I*) Polyhedral plate phytolith, identified as eudicot (MGD011). (*J*) Decorated jigsaw phytolith, likely from fruit (MGD001). (*K*) Spheroid psilate phytolith, identified as bark type (MGD001). (*L*) Damaged Triticeae starch in brightfield, with *Inset* showing an absence of birefringence in cross-polarized light (MGD010). (Scale bars, 20 μm.)

Finally, we identified the turmeric protein turmerin, supported by three peptides and four PSMs, in a single individual from Megiddo (MGD018). Turmerin is an α-amylase/trypsin inhibitor that belongs to the leguminous kunitz-type serine inhibitor family of proteases. It has known antioxidant activity (78) and also plays a role in plant defense. Although making up only 0.1% of the dry-weight of turmeric, it is one of the most stable turmeric proteins, being resistant to heat, digestive enzymes, and UV irradiation (78).

## Discussion

Among the identified dietary taxa, wheat and date palm were expected finds, as wheat has been a staple crop in the Levant since the seventh millennium BCE (79) and date palm fruits have been consumed and traded since at least the fifth millennium BCE (54, 57, 80). Moreover, date palm seeds and leaves have been previously recovered from Late Bronze II burials at Megiddo (81). Microfossil remains of both crops have been previously identified in the Bronze Age Levant (82–84), and here we identified both wheat and date palm phytoliths, as well as wheat proteins, including a major wheat gluten.

Sesame and millet, although not unexpected, are important new finds. These nonlocal domesticates from South and East Asia, respectively, spread to West Asia during the Bronze Age (85), but their arrival in the Levant is less well understood. To date, the oldest remains of sesame seeds (*Sesamum indicum*) have been found at Harappan sites in the Indus Valley (2500 to 2000 BCE), but charred and desiccated seeds have also been sporadically recovered at sites in the Near East since the late third millennium BCE (86, 87). The Akkadian word "šamaš-šammû," which refers to an oil plant (possibly sesame), appears in cuneiform texts from 2400 BCE onwards (87–89), and although questioned in the past, the identification of sesame seeds in the tomb of Tutankhamun (14th century BCE) is now considered credible (90). We found robust evidence for multiple *Sesamum* proteins in individuals at both Megiddo and Tel Erani, suggesting that by the second millennium BCE, sesame had become a staple oil-bearing crop in the Levant.

We identified Panicoideae phytoliths consistent with dietary millets in a single individual at Tel Erani. Although not identifiable below the taxonomic level of subfamily, the Eurasian grasses *Setaria* and *Panicum* and the African grasses *Sorghum* and *Pennisetum* are possible candidates. Among these, *Sorghum* and *Pennisetum* are unlikely, as there is no evidence for the dispersal of sorghum into the Levant prior to the medieval period (54), and although African pearl millet (*Pennisetum glaucum*) had spread from East Africa to South Asia by the mid-second millennium BCE (91, 92), there is no evidence for its cultivation in the Near East. In contrast, Asian broomcorn (*Panicum miliaceum*) and foxtail (*Setaria italica*) millets are known to have reached western Eurasia via Central Asia (93, 94) by the second or possibly third millennium BCE (54, 95), although likely no earlier based on recent radiocarbon dating and collagen stable isotope studies (96). Within the Near East, current macrobotanical evidence suggests that *S. italica* and *P. miliaceum* millets functioned as minor crops from the first millennium onwards (54, 97–100), and this is consistent with our identification of the panicoid phytoliths at Iron Age Tel Erani.

In contrast to the crops above, the consumption of soybean (*Glycine*), probable banana (*Musa*), and turmeric (*Curcuma longa*) were unexpected finds. Soybean cultivation was unknown in this region before the 20th century CE and, like millet, its domestication center was near the Yellow River in Central China, where it was cultivated as early as 7000 to 6500 BCE (101). However, soybean, like sesame, is a major oil plant, and its oil could have been transported over long distances. Exotic oils are frequently mentioned in Old Babylonian, Akkadian, and Egyptian texts (102–104). However, many of the plants from which they derive remain untranslated or unknown; for example, it is still disputed whether Egyptian "baq oil" refers to olive oil, moringa oil, or to the oil of another plant (103). There was high

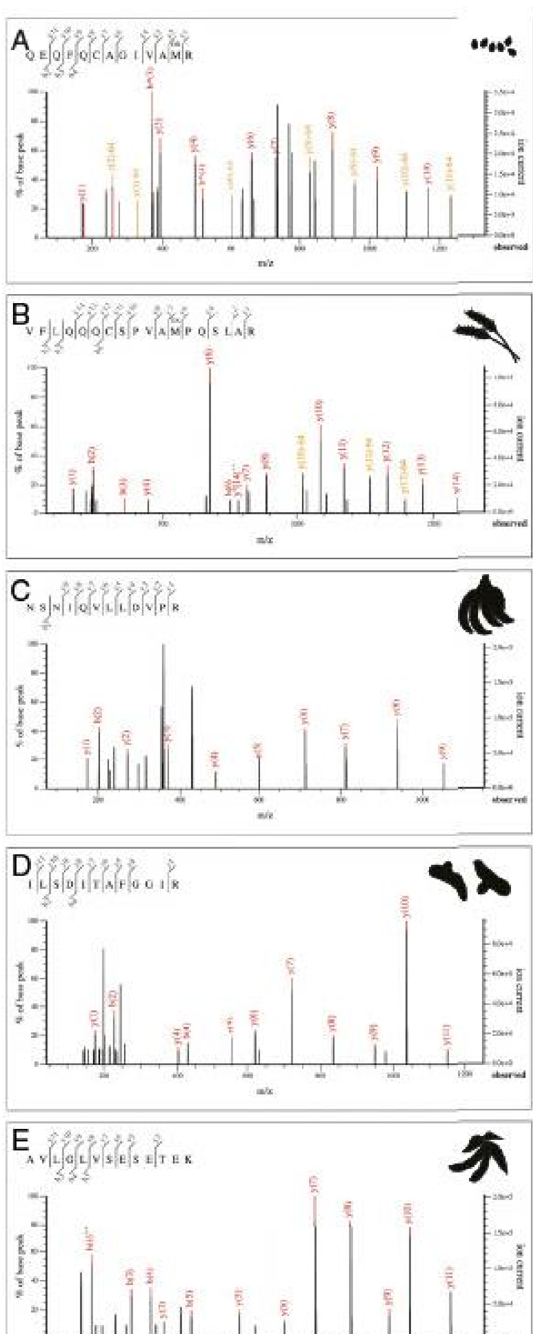



Fig. 4. Representative MS/MS spectra of selected dietary peptides. (*A*) *Sesamum*, 11S globulin protein (MGD011). (*B*) *Triticum*/*Aegilops*, LMW glutenin (MGD011). (*C*) *Musa*, β-1,3 glucanase (ERA017). (*D*) *Curcuma*, turmerin (MGD018). (*E*) *Glycine*, sucrose-binding protein (MGD018).

demand for oils of different flavor and origin in ancient Mesopotamia and Egypt, where such oils played vital roles in cuisine, medication, bodycare, and illumination. Furthermore, they were part of a broad spectrum of daily and ritual practices, where they were used, for example, to anoint objects and embalm the dead (103–106). Recent X-ray computed tomography imaging of archaeological soybeans indicates that oil content was a major target of selection during domestication, and cultivars with high oil content were prevalent in China by approximately 2000 BCE (107). The relative scarcity of evidence for soybean in the archaeological record might be explained by its predominant use as an oil combined with a lack of organic residue analysis on relevant material and/or the difficulty of differentiating plant oils through lipid analysis. Proteomics has been previously shown to be a powerful tool for identifying archaeological oils and fats (108), as plant oils and rendered animal fats produced using preindustrial techniques generally contain residual proteins. By identifying strong evidence of soybean seed proteins in dental calculus, we confirm that Megiddo individual MGD018, who was also buried with other exotic grave goods (Table 1 and *SI Appendix*), likely had access to soybean oil, and we further demonstrate the utility of proteomics as a method for identifying exotic oils in the Near East that otherwise leave few archaeological traces.

Banana (*Musa*) was domesticated during the fifth millennium BCE in New Guinea (109) and by the first millennium BCE had dispersed under human cultivation as far west as Cameroon in West Africa (110, 111). However, reconstructing the intervening cultivation and trade of bananas has proven particularly difficult to trace through the archaeological record. Banana fruit is highly perishable and domesticated bananas are seedless (reproducing instead by cuttings), and thus there is a strong bias against the recovery of banana macrobotanical remains. Phytoliths represent the principle method employed to trace early banana use, but banana phytolith recovery from archaeological contexts is generally low, perhaps due to the limited number of phytoliths produced by the plant, the lack of phytoliths in the consumed mesocarp, and other taphonomic factors (112, 113). Direct evidence of banana, in the form of phytoliths, has been found at three Indus sites dating to the late third millennium (114–117), placing the crop in South Asia by that date. The earliest reported archaeobotanical evidence for banana in the Near East is desiccated fruit pulp recovered from a vessel in an Egyptian 18th Dynasty tomb (15th ct. BCE), but the identification is highly contested (118, 119). Other scholars have argued for an introduction of banana in the Arabian peninsula prior to the ninth century BCE (120), but the first secure identification of *Musa* in the Near East consists of a banana leaf found in an Egyptian tomb at Antinoë dating to the fifth century CE (121), long after the crop had already spread further west to become a staple in West Africa (110, 111). The prehistoric spread and use of banana in South and West Asia remains poorly known, in large part due to preservation biases. Our identification of a major banana fruit-ripening protein in the dental calculus of individual ERA017 at Tel Erani lends support for either the banana being present in the Levant by the first millennium BCE or a mobile individual (e.g., a merchant or seafarer) who consumed banana during his lifetime in South or East Asia before being buried at Tel Erani. This identification, although only supported by two peptides, is nevertheless specific, and within the broader banana family Musaceae the peptides are consistent only with Asian bananas (*Musa*), to the exclusion of other related plants, such as the African enset (*Ensete*), also known as the Abyssinian banana. The protein itself, which is abundant in ripening fruits, is also supportive of a *Musa* fruit identification, as the fruits of the Abyssinian banana are inedible and only the starchy portions of the pseudostem and corm are consumed.

The Asian tropical plant family Zingiberaceae contains numerous economically useful plants used for food, spices, medicines, dyes, and perfumes (122, 123). Among these, the rhizomes of the genus *Curcuma* are consumed as foods in South and Southeast Asia, of which domesticated turmeric (*C. longa*) is among the most important and widely used. Turmeric starch grains have been identified in both cattle dental calculus and pottery at the Harappan site of Farmana, dating to between 2600 and 2200 BCE (124–126). Turmeric has multiple uses as both a spice and a cloth dye (72), and early medical texts in China and India also report its use as a medicinal (127). Within the Near East, the earliest references to turmeric appear during the seventh century BCE in Assyrian cuneiform medical texts from Ashurbanipal's library at Nineveh (128), but no archaeological evidence has been found prior to the Islamic period during the 11th to 13th centuries CE (1). Our identification of *Curcuma* protein at the site of Megiddo suggests that it was already present or accessible to individuals in the Levant as early as the mid-second millennium BCE. Interestingly, turmeric protein was identified at Megiddo in the same individual (MGD018) whose dental calculus contained soybean protein. This individual was buried in a wealthy collective tomb containing exotic goods (Table 1 and *SI Appendix*), suggesting that this individual was either well connected with trading activities or may have even been a merchant or trader themself. As such, the individual may have consumed foods seasoned with turmeric or prepared with soy oil in the Levant, in South Asia, or elsewhere.

Overall, the plant microremains and proteins identified in dental calculus from Megiddo and Tel Erani point toward the existence of a dynamic and complex exchange network connecting the Mediterranean with South Asia during the second millennium BCE that outlived the dramatic socio-political transformation and associated shift from centrally organized trade during the Bronze Age to diverse small-scale trade entrepreneurship from the Early Iron Age onwards. Historically, archaeologists and historians in the Near East have relied on texts, iconography, and macrobotanical remains as their primary sources of information in reconstructing the region's cuisine and trade connections. However, these approaches alone can be limiting. Although the text base is rich, including both cuneiform texts and papyri, many of the attested plant-related terms cannot yet be translated or fully understood, despite being transliterated. Iconography, while providing vivid images of the past, can lack the botanical detail necessary to make conclusive identifications, and macrobotanical remains are generally strongly biased toward grains and other seeds, especially those that have been carbonized. While important, these methods can miss oils, fruits, and spices that are unlikely to carbonize and which may have been traded in small quantities or in already processed powdered or liquid forms.

Here we demonstrate the utility of combining microscopic and molecular techniques to reveal a broader range of foodstuffs that include both staple grains, as well as oils, fruits, and spices that otherwise leave behind little macrobotanical evidence. In particular, we found that plant phytoliths primarily reflected bulk dietary items, and were dominated by high levels of grass phytoliths, most likely deriving from wheat, but also including probable millets. We also identified palm phytoliths, almost certainly from date fruit. All of these resources are monocots, reflecting their tendency to be well represented by phytoliths. In contrast, protein analysis revealed a wider diversity of foods and is better at identifying plants with low levels of silicification and therefore few phytoliths, especially if they are protein-rich. Although the modes of protein preservation and incorporation in ancient dental calculus are still under study, most of the proteins we identified were either protease inhibitors or belonged to the seed storage protein superfamily. Both types of proteins are highly stable against proteolysis and thermal processing, traits that may increase their likelihood of survival in archaeological

ANTHROPOLOGY

calculus. These same traits can also contribute to their potential allergenicity (129), and many of the identified proteins are also known allergens. In addition to being resilient, several of the proteins we identified, especially seed proteins, are also highly abundant in the foods from which they originate, and similar seed storage proteins have been previously identified in Neolithic pottery residues (130) and in dental calculus from the medieval and postmedieval periods (131–133).

The identification of sesame, soybean, banana, and turmeric proteins in Megiddo and Tel Erani dental calculus points to the need to reevaluate the current evidence for the second millennium BCE Indo-Mediterranean trade. Previous suggestions of such a trade network between India and Egypt were largely ignored due to the limited physical evidence (119, 134). However, such evidence is now growing, and thus earlier identifications of plants (e.g., jasmine, nutmeg, cinnamon), both from archaeological remains and from texts, should be reconsidered. The broader body of evidence for exotic goods, which also includes zebu cattle, chickens, citron, melon, cloves, millet, vanillin, peppercorns, monkeys, and beetles, points to a pattern of established trade. The individuals from Megiddo, tomb 100 (MGD017, MGD018) not only had access to exotic food, but were also buried with precious grave goods (Dataset S2), which might indicate a higher-status position or the collective burial of members of a trading house. Such traders and travelers may have transported cargoes of animals, spices, dried fruits, oils, and perfumes via different routes, either overland through Iran and the Central Asian oases or by sea across the Indian Ocean to either the Red Sea or Persian Gulf, or both.

This study highlights the potential of microscopic and molecular methods to reveal elements of trade and cuisine that otherwise leave few archaeological traces. As detection methods continue to improve, there will likely need to be a fundamental reconsideration of the dimensions and complexities of Bronze Age trans-Eurasian trade. Although named for a metal that is highly visible in the archaeological record, the process of Bronzization was likely a much broader phenomenon that also linked cuisines and economies across Eurasia.

## Materials and Methods

**Excavation.** All individuals analyzed in this study were excavated and documented in their archaeological context from the sites of Megiddo and Tel Erani within the last decade (Dataset S1). An archaeological and anthropological overview of each burial is provided in *SI Appendix*. Burials have been dated using radiometric methods and/or associated grave goods.

**Sampling.** Dental calculus sampling was performed in a clean laboratory at the Tel Aviv University Megiddo excavation archives and at the Tel Erani excavation storage facility of the Israel Antiquity Authorities. Nitrile gloves were worn during collection, and calculus was sampled using dental curettes that were replaced or cleaned with isopropanol between samples. Calculus was collected onto weighing paper or aluminum foil and packaged individually. Samples were received at the Max Planck Institute for the Science of Human History ancient proteomics laboratory, where they were weighed and subsampled prior to protein extraction. Approximately 5 to 10 mg of dental calculus was used for protein analysis, and 0.5 to 2 mg of calculus was subsampled and transferred into 1.5-mg Eppendorf tubes for analysis by microscopy. These samples, along with nine pellets remaining after protein extraction were transported to the Max Planck Institute for Evolutionary Anthropology for analysis of microremains.

**Microremains.** Prior to analysis, we first tested for the presence of surface contaminants in a subset of samples ($n = 2$; MGD002, MGD016) by performing a staged decalcification to compare the relative abundance and composition of surface and interior microremains (135, 136). We found surface contamination to be negligible (*SI Appendix*, Fig. S11), and so proceeded to process the remaining samples.

To extract the microremains we added ~1 mL of 0.5 M EDTA to decalcify our preweighed dental calculus samples that were in 1.5-mL Eppendorf tubes under a Bio Air Aura Mini laminar flow in the Department of Primatology at the Max Planck Institute for Evolutionary Anthropology, Leipzig. Samples were left in EDTA until decalcification was complete, which varied from a few hours to a few days. In two samples (MGD002 and MGD016), we performed a predecalcification to separate microremains on the outside of the calculus pieces from the interior (staged decalcification) (*SI Appendix*). The samples were then centrifuged at 2,000 × *g* for 10 min (Roth Minicentrifuge) and EDTA was removed from the samples by pipetting the supernatant. This process was repeated three times. Then 25% glycerine for mounting was added to the tube. In addition to analyzing whole calculus pieces, cellular debris pellets left over after protein extraction (see next section) were also analyzed for ERA005, ERA017, ERA023, MGD001, MGD002, MGD009, MGD010, MGD011, and MGD017. For these pellets, 100 µL of 25% glycerine solution was slowly added to the tubes to avoid spillage loss due to foaming of residual SDS from protein extraction. For all samples, 20 µL of each sample were mounted on glass slides with 18 × 18- or 22 × 22-mm coverslips depending on volume. The mounting was performed in a laminar flow hood and examined under brightfield and cross-polarized light on a Zeiss Axioscope microscope at 400x magnification (Num. Aperture = 0.95). Microremains were analyzed by examining the whole slide and any encountered microremains were photographed, described, and documented using the procedures described in *SI Appendix, Evidence of Long-Distance Trade in the Ancient World, and Procedures for Microremains Analysis*.

**Proteomics.** Protein extractions were performed on 14 dental calculus samples using a filter-aided sample preparation protocol modified for ancient proteins (see published protocol at https://www.protocols.io/view/ancient-proteins-extraction-protocol-7vwhn7e). Samples were extracted and digested alongside negative extraction blanks in order to monitor for potential laboratory contamination. Cellular debris pellets left over after protein extraction were set aside for microscopic examination of microremains (see above section). Extracted peptides were analyzed by LC-MS/MS using a Q-Exactive mass spectrometer (Thermo Scientific) coupled to an ACQUITY UPLC M-Class system (Waters) at the Functional Genomics Center Zurich of the University/Eidgenössische Technische Hochschule Zurich (*SI Appendix*). Injection blanks were also run between each sample in order to identify and reduce potential carryover across samples. MS/MS spectra were converted to Mascot generic files by MSConvert v3.0.11781 using the 100 most-intense peaks in each spectra. All MS/MS samples were analyzed using Mascot (Matrix Science; v2.6.0). Mascot was set up to search the SwissProt Release 2019_08 database (560,823 entries) and Uniprot Trembl 2017_07 (88,032,926 sequences) assuming the digestion enzyme trypsin, with an automatic decoy option. Mascot was searched with a fragment ion mass tolerance of 0.050 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Deamidation of asparagine and glutamine and oxidation of methionine and proline were specified in Mascot as variable modifications. All protein identifications were established at a protein false-discovery rate of less than 3.0% and peptide false-discovery rate of less than 1.0% using the Protein Prophet algorithm (137) implemented in the software program Scaffold (version Scaffold_4.8.9; Proteome Software). Dietary proteins were filtered at a minimum of 95% protein identification probability. Additionally, only proteins with a minimum of two unique peptides with at least one species-specific peptide were accepted (Dataset S5). Proteins that contained similar peptides that could not be differentiated on the basis of MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. No dietary proteins were found in the extraction blanks, which contained only reagents, such as porcine trypsin, and known laboratory contaminants, such as collagen, keratin, or bacterial proteins. Representative MS/MS spectra of selected dietary proteins are provided in *SI Appendix*, Fig. S10. A complete list of all identified dietary spectra is provided in Dataset S5.

**Data Availability.** Protein spectra have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (https://www.ebi.ac.uk/pride) under the dataset identifier PXD021498.

**ACKNOWLEDGMENTS.** This study was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research innovation programme (ERC-2015-StG 678901-Food-Transforms) as part of P.W.S.'s ERC Starting Grant project "FoodTransforms: Transformations of Food in the Eastern Mediterranean Late Bronze Age." Work at Megiddo is currently supported by the Shmunis Family Foundation, the Dan David Foundation, Jacques Chahine, Mark Weissman, as well as Norman and Vivian Belmonte. We thank Zeev Herzog and Lily Avitz Singer for giving us the permission to publish the pottery sherd photograph and for supplying information as to its context; and Chen Tao and Robert Spengler for helpful comments on the manuscript.

1. M. Van der Veen, J. Morales, The Roman and Islamic spice trade: New archaeological evidence. *J. Ethnopharmacol.* 167, 54–63 (2015).
2. P. Freedman, *Out of the East: Spices and the Medieval Imagination* (Yale University Press, 2008).
3. F. Wood, *The Silk Road: Two Thousand Years in the Heart of Asia* (University of California Press, 2002).
4. H. Vandkilde, Bronzization: The Bronze Age as pre-modern globalization. *Praehist. Z.* 91, 103–123 (2016).
5. K. M. Linduff, *Metallurgy in Ancient Eastern Eurasia from the Urals to the Yellow River* (Edwin Mellen Press, 2004).
6. K. Kristiansen, P. Suchowska-Ducke, Connected histories: The dynamics of Bronze Age interaction and trade 1500–1100 BC. *Proc. Prehist. Soc.* 81, 361–392 (2015).
7. E. Pernicka, "The expansion of tin bronze in the 3rd millennium" [in German] in *Mensch und Umwelt in der Bronzezeit Europas*, B. Hänsel, Ed. (Oetker-Voges, 1998), pp. 135–147.
8. J. M. Kenoyer, "Indus and Mesopotamian trade networks: New insights from shell and carnelian artifacts" in *Intercultural Relations between South and Southwest Asia: Studies in Commemoration of E.C.L. During Caspers (1934–1996)*, E. Olijdam, R. H. Spoor, Eds. (Archaeopress, 2008), pp. 19–28.
9. D. Q. Fuller, N. Boivin, T. Hoogervorst, R. Allaby, Across the Indian Ocean: The prehistoric movement of plants and animals. *Antiquity* 85, 544–558 (2011).
10. N. Boivin, A. Crowther, M. Prendergast, D. Q. Fuller, Indian Ocean food globalisation and Africa. *Afr. Archaeol. Rev.* 31, 547–581 (2014).
11. T. D. Hall, *Comparing Globalizations: Historical and World-Systems Approaches* (Springer, 2017).
12. S. C. Chew, *The Southeast Asia Connection: Trade and Polities in the Eurasian World Economy, 500 BC–AD 500* (Berghahn Books, 2018).
13. L. Rahmstorf, "The use of bronze objects in the 3rd millennium BC—A survey between Atlantic and Indus" in *Appropriating Innovations. Entangled Knowledge in Eurasia, 5000–1500 BCE*, J. Maran, P. Stockhammer, Eds. (Oxbow Books, 2017), pp. 184–210.
14. S. C. Chew, *The Recurring Dark Ages: Ecological Stress, Climate Changes, and System Transformation* (Altamira Press, 2007).
15. E. H. Cline, Sailing the Wine-Dark Sea: International Trade and the Late-Bronze Age Aegean. *Tempus Reparatum* (1994).
16. C. M. Monroe, *Scales of Fate: Trade, Tradition, and Transformation in the Eastern Mediterranean, Ca. 1350-1175 BCE* (Ugarit-Verlag, 2009).
17. J. Aruz, S. B. Graff, Y. Rakic, *Cultures in Contact: From Mesopotamia to the Mediterranean in the Second Millennium B.C* (Metropolitan Museum of Art, 2013).
18. P. R. S. Moorey, *Ancient Mesopotamian Materials and Industries: The Archaeological Evidence* (Eisenbrauns, 1999).
19. G. Barjamovic, "Interlocking commercial networks and the infrastructure of trade in Western Asia during the Bronze Age" in *Trade and Civilisation: Economic Networks and Cultural Ties from Prehistory to the Early Modern Era*, K. Kristiansen, T. Lindkvist, J. Myerdal, Eds. (Cambridge University Press, 2018), pp. 113–142.
20. G. Barjamovic, "The geography of trade. Assyrian colonies in Anatolia c. 1975-1725 BC and the study of early interregional networks of exchange" in *Anatolia and the Jazira during the Old Assyrian Period*, J. G. Derksen, Ed. (Nederlands Instituutvoor het Nabije Osten, 2008), pp. 87–100.
21. R. Cohen, R. Westbrook, Eds., *Amarna Diplomacy: The Beginnings of International Relations* (The Johns Hopkins University Press, 2000).
22. S. Baumann, *Treasuries: Their Decoration and Spatial Conception in Egyptian Temples of the Greco-Roman Period* [in German] (Harrassowitz Verlag, 2018).
23. R. Herzog, *Punt* (Augustin, 1968).
24. D. K. Chakrabarti, *The External Trade of Indus Civilization* (Munshiram Manoharlal Publishers, 1990).
25. H. E. W. Crawford, Mesopotamia's invisible exports in the third millennium BC. *World Archaeol.* 5, 232–241 (1973).
26. G. L. Possehl, Seafaring merchants of Meluhha. *South Asian Archaeol.* 1995, 87–100 (1997).
27. S. Parpola, A. Parpola, R. H. Brunswig, The Meluhha Village: Evidence of acculturation of Harappan traders in late third millennium Mesopotamia? *J. Econ. Soc. Hist. Orie.* 20, 129–165 (1977).
28. G. L. Possehl, Shu-ilishu's cylinder seal. *Expedition* 48, 42–43 (2006).
29. C. J. Gadd, *Seals of Ancient Indian Style Found at Ur* (H. Milford, 1933).
30. J. Reade, "The Indus-Mesopotamian relationship reconsidered" in *Intercultural Relations between South and Southwest Asia: Studies in Commemoration of E.C.L. during Caspers (1934-1996)*, E. Olijdam, R. H. Spoor, Eds. (Archaeopress, Oxford, 2008), pp. 12–18.
31. J. F. Jarrige, J. F. Jarrige, G. Quivron, *The Forgotten Cities of the Indus* [in French] (Archéologie du Pakistan, 1988).
32. D. S. Reese, The trade of Indo-Pacific shells into the Mediterranean Basin and Europe. *Oxf. J. Archaeol.* 10, 159–196 (1991).
33. J. McIntosh, *The Ancient Indus Valley: New Perspectives* (ABC-CLIO, 2008).
34. S. K. Stein, *The Sea in World History: Exploration, Travel, and Trade [2 volumes]* (ABC-CLIO, 2017).
35. E. R. Arnold *et al.*, Isotopic evidence for early trade in animals between Old Kingdom Egypt and Canaan. *PLoS One* 11, e0157650 (2016).
36. M. Meiri, P. W. Stockhammer, P. Morgenstern, J. Maran, Mobility and trade in Mediterranean antiquity: Evidence for an "Italian connection" in Mycenaean Greece revealed by ancient DNA of livestock. *J. Archaeol. Sci. Rep.* 23, 98–103 (2019).
37. M. Meiri *et al.*, Eastern Mediterranean mobility in the Bronze and Early Iron Ages: Inferences from ancient DNA of pigs and cattle. *Sci. Rep.* 7, 701 (2017).
38. S. Chen *et al.*, Zebu cattle are an exclusive legacy of the South Asia Neolithic. *Mol. Biol. Evol.* 27, 1–6 (2010).
39. R. Matthews, Zebu: Harbingers of doom in Bronze Age Western Asia? *Antiquity* 76, 438–446 (2002).
40. M. P. Verdugo *et al.*, Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* 365, 173–176 (2019).
41. M. N. Pareja *et al.*, A new identification of the monkeys depicted in a Bronze Age wall painting from Akrotiri, Thera. *Primates* 61, 159–168 (2020).
42. E. Panagiotakopulu, Insect remains from the collections in the Egyptian Museum of Turin. *Archaeometry* 45, 355–362 (2003).
43. L. Perry-Gal, A. Erlich, A. Gilboa, G. Bar-Oz, Earliest economic exploitation of chicken outside East Asia: Evidence from the Hellenistic Southern levant. *Proc. Natl. Acad. Sci. U.S.A.* 112, 9849–9854 (2015).
44. G. Fritz, M. Nesbitt, "Laboratory analysis and identification of plant macroremains" in *Method and Theory in Paleoethnobotany*, J. M. Marston, J. D. A. Guedes, C. Warinner, Eds. (University Press of Colorado Boulder, 2015), pp. 115–145.
45. X. Zhou *et al.*, 5,200-year-old cereal grains from the eastern Altai Mountains redate the trans-Eurasian crop exchange. *Nat. Plants* 6, 78–87 (2020).
46. X. Liu, H. V. Hunt, M. K. Jones, River valleys and foothills: Changing archaeological perceptions of North China's earliest farms. *Antiquity* 83, 82–95 (2009).
47. N. Boivin, D. Q. Fuller, Shell middens, ships and seeds: Exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula. *J. World Prehist.* 22, 113–180 (2009).
48. D. Langgut, The Citrus Route revealed: From Southeast Asia into the Mediterranean. *HortScience* 52, 814–822 (2017).
49. G. A. Wu *et al.*, Genomics of the origin and evolution of citrus. *Nature* 554, 311–316 (2018).
50. D. Langgut, Y. Gadot, N. Porat, O. Lipschits, Fossil pollen reveals the secrets of the Royal Persian Garden at Ramat Rahel, Jerusalem. *Palynology* 37, 115–129 (2013).
51. M. Van der Veen, "The botanical evidence" in *Survey and Excavations at Mons Claudianus 1987-1993*, V. A. Maxfield, D. P. S. Peacock, Eds. (Institut Français d'Archéologie Orientale du Cairo, 2001), pp. 174–247.
52. M. Van der Veen, H. Tabinor, "Food, fodder and fuel at Mons Porphyrites: The botanical evidence" in *The Roman Imperial Quarries: Survey and Excavation at Mons Porphyrites 1994-1998*, D. Peacock, V. Maxfield, Eds. (The Egypt Exploration Society, London, 2007), pp. 84–156.
53. H. Hjelmqvist, "Some economic plants and weeds from the Bronze Age of Cyprus" in *Hala Sultan Tekke 5, Studies in Mediterranean Archaeology*, U. Öbrink, Ed. (Åströms Förlag, 1979), vol. 45, pp. 110–133.
54. D. Zohary, M. Hopf, E. Weiss, *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford University Press, 2012).
55. P. Sebastian, H. Schaefer, I. R. Telford, S. S. Renner, Cucumber (Cucumis sativus) and melon (C. melo) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14269–14273 (2010).
56. S. Muthukumaran, "An ecology of trade: Tropical Asian cultivars in the ancient Middle East and the eastern Mediterranean," PhD thesis, University College London, London, United Kingdom, (2016).
57. C. de Vartavan *et al.*, *Codex of Ancient Egyptian Plant Remains* (SAIS, 2010).
58. V. Linares *et al.*, First evidence for vanillin in the Old World: Its use as mortuary offering in Middle Bronze Canaan. *J. Archaeol. Sci. Rep.* 25, 77–84 (2019).
59. A. Gilboa, D. Namdar, On the beginnings of South Asian spice trade with the Mediterranean region: A review. *Radiocarbon* 57, 265–283 (2015).
60. D. Namdar, R. Neumann, S. Weiner, "Residue analysis of chalices from the repository pit" in *Yavneh I: The Excavation of the "Temple Hill" Repository Pit and the Cult Stands*, R. Kletter, I. Ziffer, W. Zwickel, Eds. (Academic Press Fribourg, 2010), pp. 167–173.
61. R. Lichtenberg, A. C. Thuilliez, On some unusual aspects of the radiology of Ramses II. [in French] B. Mem. Soc. Anthro. Par. 3, 323–330 (1981).
62. A. Plu, "Wood and grains" in *La Momie de Ramsès: Contribution Scientifique à l'Égyptologie*, L. Balout, C. Roubet, Eds. (Éditions Recherche sur les Civilisations, 1985), pp. 166–174.
63. E. Asouti, D. Q. Fuller, *Trees and Woodlands of South India: Archaeological Perspectives* (Routledge, 2008).
64. G. Buccellati, M. K. Buccellati, Terqa: The first eight seasons. *Les Ann. Archeol. Arabes Syriennes* 32, 47–67 (1983).
65. M. Smith, "The Terqa Cloves and the Archaeology of Aroma" in *Between Syria and the Highlands; Studies in honor of Giorgio Buccellati & Marilyn Kelly-Buccellati*, S. Valentini, Ed. *et al.* (Arbor Sapientiae Editore, 2019), pp. 373–377.
66. E. Kingwell-Banham *et al.*, Spice and rice: Pepper, cloves and everyday cereal foods at the ancient port of Mantai, Sri Lanka. *Antiquity* 92, 1552–1570 (2018).
67. D. R. Piperno, *Phytoliths: A Comprehensive Guide for Archaeologists and Paleoecologists* (Rowman Altamira, 2006).
68. D. Cabanes, R. Shahack-Gross, Understanding fossil phytolith preservation: The role of partial dissolution in paleoecology and archaeology. *PLoS One* 10, e0125532 (2015).
69. R. M. Albert, F. Berna, P. Goldberg, Insights on Neanderthal fire use at Kebara Cave (Israel) through high resolution study of prehistoric combustion features: Evidence from phytoliths and thin sections. *Quat. Int.* 247, 278–293 (2012).
70. S. Rustgi, P. Shewry, F. Brouns, L. J. Deleu, J. A. Delcour, Wheat seed Proteins: Factors influencing their content, composition, and technological properties, and strategies to reduce adverse reactions. *Compr. Rev. Food Sci. Food Saf.* 18, 1751–1769 (2019).
71. Z. Wang *et al.*, New insight into the function of wheat glutenin proteins as investigated with two series of genetic mutants. *Sci. Rep.* 7, 3428 (2017).
72. K. B. Metheny, M. C. Beaudry, *Archaeology of Food: An Encyclopedia* (Rowman & Littlefield, 2015).

73. M. L. Colgrave, *Proteomics in Food Science: From Farm to Fork* (Academic Press, 2017).
74. W. J. Peumans *et al.*, Purification, characterization and structural analysis of an abundant beta-1,3-glucanase from banana fruit. *Eur. J. Biochem.* 267, 1188–1195 (2000).
75. V. Receveur-Bréchot *et al.*, Crystal structure at 1.45-A resolution of the major allergen endo-beta-1,3-glucanase of banana as a molecular basis for the latex-fruit syndrome. *Proteins* 63, 235–242 (2006).
76. G. Thulasy, A. S. Nair, Comparative analysis of β-1,3-glucanase protein sequences and structure in two banana cultivars from Kerala. *J. Food Biochem.* 42, e12559 (2018).
77. S. Roy Choudhury, S. Roy, S. K. Singh, D. N. Sengupta, Molecular characterization and differential expression of β-1,3-glucanase during ripening in banana fruit in response to ethylene, auxin, ABA, wounding, cold and light-dark cycles. *Plant Cell Rep.* 29, 813–828 (2010).
78. L. Srinivas, V. K. Shalini, M. Shylaja, Turmerin: A water soluble antioxidant peptide from turmeric [Curcuma longa]. *Arch. Biochem. Biophys.* 292, 617–623 (1992).
79. D. Q. Fuller, Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the Old World. *Ann. Bot.* 100, 903–924 (2007).
80. J. B. Gillet, "Botanical samples" in *Eridu*, F. Safer *et al.*, Eds. (Ministry of Culture and Information, State Organization of Antiquities and Heritage, Baghdad, 1981), pp. 317–318.
81. N. Liphschitz, "Wood remains" in *Megiddo V: The 2004–2008 Seasons Monograph Series No. 31*, I. Finkelstein *et al.*, Eds. (Institute of Archaeology, Tel Aviv University, 2013), pp. 1220–1236.
82. C. E. White, M. S. Chesson, R. Thomas Schaub, A recipe for disaster: Emerging urbanism and unsustainable plant economies at Early Bronze Age Ras an-Numayra, Jordan. *Antiquity* 88, 363–377 (2014).
83. P. L. Fall, L. Lines, S. E. Falconer, Seeds of civilization: Bronze Age rural economy and ecology in the Southern Levant. *Ann. Assoc. Am. Geogr.* 88, 107–125 (1998).
84. A. Damick, The first identification of Phoenix dactylifera (date palm) from Early Bronze Age Lebanon. *Veg. Hist. Archaeobot.* 28, 583–589 (2019).
85. H. V. Hunt *et al.*, Millets across Eurasia: Chronology and context of early records of the genera Panicum and Setaria from archaeological sites in the Old World. *Veg. Hist. Archaeobot.* 17, 5–18 (2008).
86. V. Zech-Matterne, M. Tengberg, W. Van Andringa, Sesamum indicum L. (sesame) in 2nd century BC Pompeii, southwest Italy, and a review of early sesame finds in Asia and Europe. *Veg. Hist. Archaeobot.* 24, 673–681 (2015).
87. D. Bedigian, History and lore of sesame in Southwest Asia. *Econ. Bot.* 58, 329–353 (2004).
88. D. Bedigian, J. R. Harlan, Evidence for cultivation of sesame in the ancient world. *Econ. Bot.* 40, 137–154 (1986).
89. M. A. Powell, Epistemology and Sumerian agriculture: The strange case of sesame and linseed *Aula Orientalis* 9, 155–164 (1991).
90. D. Q. Fuller, Further evidence on the prehistory of sesame. *Asian Agrihist.* 7, 127–137 (2003).
91. K. Manning, R. Pelling, T. Higham, J.-L. Schwenniger, D. Q. Fuller, 4500-Year old domesticated pearl millet (Pennisetum glaucum) from the Tilemsi Valley, Mali: New insights into an alternative cereal domestication pathway. *J. Archaeol. Sci.* 38, 312–322 (2011).
92. D. Q. Fuller, "African crops in prehistoric South Asia: a critical review" in *Food, Fuel and Fields: Progress in African Archaeobotany*, K. Neumann, A. Butler, S. Kahlheber, Eds. (Africa Prehistorica 15, 2003), pp. 239–271.
93. N. Miller, Rainfall seasonality and the spread of millet cultivation in Eurasia. Iranian. *J. Archaeol. Stud.* 5, 1–10 (2015).
94. N. F. Miller, R. N. Spengler, M. Frachetti, Millet cultivation across Eurasia: Origins, spread, and the influence of seasonal climate. *Holocene* 26, 1566–1575 (2016).
95. G. Motuzaite-Matuzeviciute, R. A. Staff, H. V. Hunt, X. Liu, M. K. Jones, The early chronology of broomcorn millet (Panicum miliaceum) in Europe. *Antiquity* 87, 1073–1085 (2013).
96. M. A. Tafuri, O. E. Craig, A. Canci, Stable isotope evidence for the consumption of millet and other plants in Bronze Age Italy. *Am. J. Phys. Anthropol.* 139, 146–153 (2009).
97. L. Kubiak-Martens, "Plant remains from Tell Ashara (Terqa) and Tell Masaikh in the Middle Euphrates, south-eastern Syria" in *Akh Purattim*, J. Magueron, O. Rouault, P. Butterlin, P. Lombard, Eds. (Maison de l'Orient et de la Méditerranée & Ministère des Affaires étrangères et du Développement International, 2015), vol. 3, pp. 423–442.
98. A. Soltysiak, H. Schutkowski, Stable isotopic evidence for land use patterns in the Middle Euphrates Valley, Syria. *Am. J. Phys. Anthropol.* 166, 861–874 (2018).
99. M. Nesbitt, G. D. Summers, Some recent discoveries of millet (Panicum miliaceum and Setaria italica) at excavations in Turkey and Iran. *Anatol. Stud.* 38, 85–97 (1988).
100. M. Nesbitt, J. Bates, G. Hillman, S. Mitchell, *The Archaeobotany of Aşvan: Environment and Cultivation in Eastern Anatolia from the Chalcolithic to the Medieval Period* (British Institute at Ankara, 2017).
101. G.-A. Lee, G. W. Crawford, L. Liu, Y. Sasaki, X. Chen, Archaeological soybean (Glycine max) in East Asia: Does size matter? *PLoS One* 6, e26720 (2011).
102. R. Middeke-Conlin, *The Scents of Larsa: A Study of the Aromatics Industry in an Old Babylonian Kingdom* (Cuneiform Digital Library Journal, 2014), vol. 1, pp. 1–53.
103. B. Koura, *The "7 Sacred Oils" and Other Oil and Fat Names. A Lexicographical Study of the Names of Oils, Fats and Ointments Among the Ancient Egyptians from Early Dynastic Until the Beginning of the Ptolemaic Period (From 3000 BC–ca. 305 BC)* [in German] (Shaker, 1999).
104. M. Görg, Oils from abroad [in German]. *Studien Altägyptischen Kultur* 11, 219–226 (1984).
105. D. Eitam, M. Heltzer, Eds., *Olive Oil in Antiquity: Israel and Neighboring Countries, from the Neolithic to the Early Arab Period* (Sargon, 1996).
106. M. Serpico, "Natural product technology in New Kingdom Egypt" in *Invention and Innovation: The Social Context of Technological Change 2: Egypt, the Aegean and the Near East, 1650-1150 B.C*, J. Bourriau, J. Phillips, Eds. (Oxbow Books, 2004), pp. 96–120.
107. Y. Zong *et al.*, Selection for oil content during soybean domestication revealed by X-ray tomography of ancient beans. *Sci. Rep.* 7, 43595 (2017).
108. A. Shevchenko *et al.*, Open sesame: Identification of sesame oil and oil soot ink in organic deposits of Tang Dynasty lamps from Astana necropolis in China. *PLoS One* 12, e0158636 (2017).
109. X. Perrier *et al.*, Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11311–11318 (2011).
110. C. M. Mbida, W. Van Neer, H. Doutrelepont, L. Vrydaghs, Evidence for banana cultivation and animal husbandry during the first millennium BC in the forest of Southern Cameroon. *J. Archaeol. Sci.* 27, 151–162 (2000).
111. C. M. Mindzie, L. Vrydaghs, "Nkang: Early evidence for banana cultivation on the African continent" in *Encyclopedia of Global Archaeology*, C. Smith, Ed. (Springer, New York, 2014), pp. 5293–5295.
112. R. C. Power, T. Güldemann, A. Crowther, N. Boivin, Asian crop dispersal in Africa and Late Holocene human adaptation to tropical environments. *J. World Prehist.* 32, 353–392 (2019).
113. M. J. Hodson, P. J. White, A. Mead, M. R. Broadley, Phylogenetic variation in the silicon composition of plants. *Ann. Bot.* 96, 1027–1046 (2005).
114. A. Lawler, European association for South Asian archaeology and art meeting. The ingredients for a 4000-year-old proto-curry. *Science* 337, 288 (2012).
115. A. Farooqui, A. S. Gaur, V. Prasad, Climate, vegetation and ecology during Harappan period: Excavations at Kanjetar and Kaj, mid-Saurashtra coast, Gujarat. *J. Archaeol. Sci.* 40, 2631–2647 (2013).
116. J. J. García-Granero, C. Lancelotti, M. Madella, P. Ajithprasad, "Plant processing activities at Loteshwar (North Gujarat, India): A micro-botanical approach" in *Man and Environment in Prehistoric and Protohistoric South Asia: New Perspectives*, A. Didier, B. Mutin, Eds. (Brepols, 2016), pp. 99–116.
117. M. Madella, "Investigating agriculture and environment in South Asia: Present and future contributions from opal phytoliths" in *Indus Ethnobiology: New Perspectives from the Field*, S. Weber, W. Belcher, Eds. (Rowman & Littlefield Publishing Group, 2003) pp. 199–249.
118. G. Bernsen, Pharmacognostic analysis of the content of a vessel from an Egyptian tomb of the 18th Dynasty. *Arch. Pharm. Chem. (Kbh)* 4, 65–70 (1976).
119. R. Germer, *Flora of Pharaonic Egypt* [in German] (Zabern, 1985).
120. K. T. S. Al-Busaidi, Banana domestication on the Arabian Peninsula: A review of their domestication history. *J. Hortic. For.* 5, 194–203 (2013).
121. W. Bonnet, Ancient plants in the necropolis of Antinoë. *J. Bot.* 19, 5–12 (1905).
122. J. R. Harlan, *Crops and Man* (American Society of Agronomy, 1975).
123. P. Sirirugsa, "Thai Zingiberaceae: Species diversity and their uses," invited lecture presented at the International Conference on Biodiversity and Bioresources: Conservation and Utilization, 23–27 November 1997, Phuket, Thailand. http://media.iupac.org/symposia/proceedings/phuket97/sirirugsa.pdf. Accessed: 15 November 2020.
124. A. Kashyap, S. Weber, Harappan plant use revealed by starch grains from Farmana, India. *Antiquity* 84, 326 (2010).
125. S. A. Weber, A. Kashyap, L. Mounce, "Archaeobotany at Farmana: New insights into Harappan plant use strategies" in *Excavations at Farmana, Indus Ethnobiology: New Perspectives from the Field*, S. A. Weber, B. Belcher, Eds. (Lexington Books, 2011), pp. 808–823.
126. A. Kashyap, S. Weber, "Starch grain analysis and experiments provide insights into Harappan cooking practices" in *Connections and Complexity*, S. A. Abraham, P. Gullapalli, T. Raczek, U. Z. Rizvi, Eds. (Routledge, 2016), pp. 177–194.
127. S. Prasad, B. Aggarwal, "Turmeric, the golden spice" in *Herbal Medicine: Biomolecular and Clinical Aspects*, I. F. F. Benzie, S. Wachtel-Galor, Eds. (CRC Press/Taylor & Francis, ed. 2, 2011) chap 13, pp. 263–288.
128. R. C. Thompson, *A Dictionary of Assyrian Botany* (British Academy, 1949).
129. J. Pekar, D. Ret, E. Untersmayr, Stability of allergens. *Mol. Immunol.* 100, 14–20 (2018).
130. J. Hendy *et al.*, Ancient proteins from ceramic vessels at Çatalhöyük West reveal the hidden cuisine of early farmers. *Nat. Commun.* 9, 4064 (2018).
131. J. Hendy *et al.*, Proteomic evidence of dietary sources in ancient dental calculus. *Proc. Biol. Sci.* 285, 20180977 (2018).
132. R. R. Jersie-Christensen *et al.*, Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. *Nat. Commun.* 9, 4744 (2018).
133. E. Cappellini *et al.*, A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* 97, 205–217 (2010).
134. A. Lüchtrath, tj-šps, the camphor tree of East Africa. [in German] *Göttinger Miszellen* 101, 43–48 (1988).
135. K. Hardy *et al.*, Dental calculus reveals potential respiratory irritants and ingestion of essential plant-based nutrients at Lower Palaeolithic Qesem Cave Israel. *Quat. Int.* 398, 129–135 (2015).
136. C. Warinner *et al.*, Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* 46, 336–344 (2014).
137. A. I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658 (2003).

# 8. Discussion

## 8.1 Dietary protein recovery and identification

### *8.1.1. Multi-disciplinary approach to dietary studies*

This thesis highlights the advantages of applying more than one analytical method to questions of diet and reconstructing the subsistence of humans in the past. The multi-method approach -combining proteomics with microremain analysis- we adopted in Manuscripts A and D proved valuable, as the methods were complementary and additive, each providing identifications not found by the other method. In other cases, the protein results provided confirmation (Manuscript B) or contradiction of (Manuscript C) previous archaeological interpretations made on the basis of zooarchaeological findings. Additionally, combining our proteomics findings with human genetic data regarding population migrations and histories allowed us to contextualize and substantially refine archaeological models of the adoption, use, and spread of subsistence technologies.

Examining our proteomics findings as a whole, we observed that bacterial and host proteins dominated our dental calculus datasets. Relatively few samples yielded unequivocal, taxonomically specific dietary remains, with the notable exception of milk proteins identified in ancient pastoralist societies and seed proteins from ancient urban environments. Dental calculus, the calcified remains of the oral microbiome, is, as expected, overwhelming bacterial in origin (Warinner et al. 2014b). Among the non-bacterial proteins identified, fibrous protein recovery was high, and the most frequently recovered proteins were collagens, mostly human-derived or nonspecific. However, because collagen and keratin are common contaminants in proteomics facilities, and due to the frequency of collagen and keratin in extraction blanks, we did not analyze those results further.

During analysis, an overarching pattern emerged to the types of proteins we could successfully recover and identify. All recovered dietary proteins fell into three main categories: milk proteins, storage proteins (especially seed storage proteins), and defense proteins (in the form of protease inhibitors and antimicrobial peptides). Animal-derived protein identifications included those from chicken eggs (ovalbumin (OVAL)) and milk from cattle, sheep, goats, and horses (BLG, A-LAC, $AS_1$-CN). Plant-derived food proteins identifications included seed storage proteins from sesame, soybean, and wheat (11S and 7S globulins, 2S albumin, sucrose-binding protein, and glutenin) and plant defense proteins from banana, turmeric, and wheat (β-1,3 glucanase, turmerin, and alpha-amylase inhibitor (AAI)).

In manuscript A, "*Relief Food Subsistence Revealed by Microparticle and Proteomics Analyses of Dental Calculus from Victims of the Great Irish Famine*", our proteomic and microremain results gave distinctive but complementary results. Although plant proteins were recovered, we could not identify the plant proteins to a level of taxonomic specificity beyond the category of Viridiplantae or "green plants." However, we were able to identify egg and milk proteins to the genus level; these foods are not detectable using microremains approaches. In contrast, the microremain results provided good evidence of the consumption of various classes of plants, including cereals grains such as maize, oats, and wheat. This difference in recovery can be partly explained because cereals, being grasses, produce abundant plant microremains in the form of phytoliths. Our results both confirmed and augmented historical data. Our milk findings confirm historical records of milk being supplied to workhouses, but our egg protein (OVAL) identification was surprising because eggs were rarely mentioned in historical documents about food items available in the

workhouses. However, our recovered proteins cannot be narrowed down to a specific time within an individual's life, and dietary proteins could have easily been incorporated prior to the Irish famine. Although we could not specifically identify the plants consumed using proteomics, the microremains results confirmed that they were expected taxa included maize, a typical relief food sent from America and distributed in the workhouses, which was uncommon before the famine.

In general, our protein analysis for this study revealed relatively few dietary identifications, compared to dietary microremains, despite the relatively young age of our samples relative to other archaeological datasets. However, these samples were extracted with an earlier extraction protocol prior to the optimization described in section 1.3.1 of this thesis, and thus the limited results may reflect poor extraction recovery rather than preservation.

In manuscript B, "*The genomic origins of the Bronze Age Tarim Basin mummies*," our proteomic results were contextualized with data provided by paleogenetic analysis. Although previous studies had shown that the later burial layers (after 1600 BCE) had evidence of dairy consumption (Yang et al. 2014; Xie et al. 2016), it was unknown whether and when dairy technologies may have been introduced to the Xiaohe population through contact with neighboring groups. The genetic analysis in our manuscript provides evidence that the Tarim Basin mummies were a genetically homogenous and genetically isolated indigenous population, possibly due to their extreme desert environment. However, our results demonstrate that individuals buried in the basal layers of the Xiaohe cemetery (~2000 BCE; founding population) were already consuming dairy proteins, a technology that they likely adopted from neighboring pastoral groups prior to their colonization of the Tarim Basin. The introduction of dairying species and the adoption of dairying technology demonstrates that the indigenous populations of Xiaohe were genetically but not culturally isolated. They had engaged in cultural interactions with outside groups, but this did not contribute to gene flow.

Although we were able to identify robust evidence of consumed bovid milk proteins, we were not able to establish the presence of any taxonomically-specific plant proteins, which may in part be due to the high abundance of milk peptides in our datasets which can overwhelm the signal of other peptides. A similar pattern has been seen in other archaeological studies of dairy pastoralist populations, including Manuscript C. The limitations of non-dairy protein recovery in populations with high dairy consumption should be considered during study design and interpretation.

Our goal in manuscript C, "*Emergence and intensification of dairying in the Caucasus and Eurasian steppe,"* was to investigate dairy technology in the Northern Caucasus and surrounding regions. We aimed to establish the pre-history and prevalence of dairy technologies, including which animal species were utilized. We looked for differences in use through time, between cultural groups, and in ecotones. In a broader context, we considered the implications of Northern Caucasus dairy pastoralism on other regions. While the earliest known ruminant dairying occurred in the 7$^{th}$ and 6$^{th}$ millennia BCE in Anatolia, the lack of settlements data, and the archaeological focus on mortuary contexts have limited our knowledge about the introduction and spread of this subsistence technology into Asia and the Pontic-Caspian steppe is believed to be the source for the spread of mobile pastoralism that was critical for the heightened mobility that led to spread of pastoralism, populations, and culture from the Pontic Caspian steppe throughout Eurasia.

Although manuscript C was focused entirely on proteomic results, it was built upon previous research that included isotopic and genomic studies of the region and utilized the same samples in many cases (Wang et al. 2019; Knipper et al. 2020). These previous genomic studies allowed us to contextualize our data within the larger archaeological framework of population genetics, the large-scale movement of populations, and the spread of subsistence strategies that may have helped facilitate those migrations. An isotopically complex landscape and scant archaeological settlement data have hindered previous dietary studies of the region, although general isotopic patterns established an increase in dietary animal protein during periods when dairying was suspected of having occurred. However, dairy consumption could not be explicitly established by these previous studies. For the first time, our protein results established unequivocal evidence of dairy consumption in a region suspected to be an important area in establishing and spreading high mobility dairy pastoralism.

Importantly we were able to establish livestock-specific dairy consumption in most cases, an advantage of ancient proteomics over other methods. Importantly our results help to establish temporal and geographic boundaries of dairy consumption. Although a lack of dairy peptides cannot be used to establish a lack of consumption, we did not recover dairy proteins in any extraction blanks or from individuals within a population without archaeological evidence of domesticated dairy species, supporting previous archaeological hypotheses about the timing and adoption of dairy technologies. As in Manuscript B, our extremely high abundance of milk peptides could have hindered the recovery of other dietary proteins, including those from plant species. However, as most of our individuals did not participate in an agriculture-based subsistence, the vegetables they consumed would likely have been low in protein and less likely to be identified in the dental calculus through proteomics.

In manuscript D, "*Exotic foods reveal contact between South Asia and the Near East during the 2nd millennium BCE*", our goal was to establish the overall protein recovery expected from Bronze and Iron Age calculus from the Southern Levant, as no previous proteomic studies of calculus had occurred in the region. Importantly, our samples came from ancient sedentary urban environments and not from populations that engaged in more extreme dairy-based diets, such as those associated with mobile pastoralists, whose subsistence strategies were highly focused on domesticated animal products. We aimed to establish whether non-dairy dietary proteins would be easier to recover in this context. We also wanted to assess how complementary our proteomic method would be to microremain results, which included samples that were older than those in our Irish Famine study by thousands of years. Finally, could we recover other proteins not found with other methods or that leave few archaeological traces?

In general, the microremain and protein identifications both confirmed the consumption of the dietary staple wheat but otherwise resulted in few overlapping results. Microremain analysis established dietary plant consumption in all analyzed samples (n=16), while dietary proteins were recovered only from a subset of the studied samples (5 out of 14). Although most microremains were not taxonomically specific, many could still be narrowed down to broad classes of plants, such as grasses or cereal grains. Both analyses were able to show the presence of wheat in the same sample, although proteomics was only able to establish consumption in one individual, while the microremains provided evidence of Triticeae consumption in two additional individuals and grain consumption in eight individuals. Our two methods were also able to recover different kinds of evidence for the consumption of fruit. The microremains recovered evidence for the consumption of *Phoenix* (date palm), a

fruit that produces an abundance of morphologically distinctive phytoliths, while the proteomic analysis revealed evidence of a ripening protein from *Musa* (banana), a seedless plant that produces very few phytoliths. Similarly, proteomics provided evidence for turmeric consumption, a spice that leaves little if any archaeological traces and may remain invisible to other archaeological methods. Additionally, it recovered species-and tissue-specific evidence of seed consumption that is less likely to be recovered with any other type of molecular analysis. The specificity of these identifications allowed us to comment upon the likelihood of trade that facilitated the availability of several South Asian plant species to be incorporated in the dental calculus of Bronze Age Levantine individuals, adding to a growing body of evidence for Bronze Age Indo-Mediterranean trade.

#### ***8.1.2. Protein preservation in the archaeological record***

Although our ability to perform statistically robust experimental studies on archaeological dental calculus is limited, at present, a deeper understanding of the expected differential preservation of proteins within dental calculus can contribute to more robust studies of ancient diets. Dental calculus is not well suited for experimental studies, as it is a finite and irreplaceable material without a comparable synthetic model and high differential preservation. Therefore, in addition to the specific findings in each manuscript, we examined our entire dataset for broad patterns in protein preservation and found some protein characteristics that could potentially play a role in preserving proteins within calculus.

##### *8.1.2.a Shared characteristics from recovered dietary proteins*

Many of the recovered dietary we identified originate from foods that were known or suspected to have been consumed in the past and are still important components of human diets today. All are known to be potential allergens, and many have been well-characterized due to their potential allergenicity. We suspect the properties that induce allergenicity also aid in their survival during food preparation and their incorporation into the dental plaque of the human oral cavity. The resistance of these allergens to degradation from heat, pH changes, and proteolysis by enzymes may prolong their survival, and the presence of disordered or calcium-binding regions increases their likelihood of being involved in biomineralization events. Although our dietary proteins were all globular or intrinsically disordered, many preserved well, even relative to fibrous proteins in our datasets. In fact, in calculus samples from known pastoralist communities, globular milk proteins are often one of the highest-scoring proteins, with large numbers of duplicate peptides. Many of our identified proteins can form amyloid fibrils (AFs), insoluble self-assembled proteins, or polypeptide aggregates under specific conditions, discussed below. Although studies have not confirmed the AF-forming ability of each protein we identified, many are predicted to be capable of AF formation based on their homologous structures to other AF-forming proteins.

##### *8.1.2.b. Amyloid fibrillation may contribute to protein preservation*

Amyloid fibrils are self-assembled proteins or polypeptides that form insoluble protein aggregates, and AFs that form in the brain are associated with several disorders such as Alzheimer’s and Huntington’s diseases (Ke et al. 2020). Although the field of amyloid biology has only recently started to expand in earnest outside of pathological amyloids, a number of important studies have demonstrated just how prevalent AFs are and their important functional roles in many biological processes. The AF structure itself is extraordinarily robust, conferring soluble proteins with extreme stability, similar to collagen and AFs are highly resistant to denaturation through chemical or physical processes, even in extreme conditions (Martinez et al. 2015; Graether and Sykes 2004; Arai et al. 2019). Recently, researchers have also started to recognize the role of amyloids in protein

allergenicity and have established that many dietary allergens can form AFs in specific conditions, increasing their allergenicity as the large aggregates become stabilized and insoluble (Martínez et al. 2015; Sánchez et al. 2016). We recovered and identified these same types of dietary proteins from archaeological human dental calculus. Although AFs have been shown to play a role in bacterial biofilm formation of dental plaque (Jakubovics et al. 2021), no studies have yet looked for the presence of AFs within mineralized dental calculus.

How these fibrils would form and incorporate into the dental calculus is currently unknown. However, it likely involves the same protein interactions involved in binding bacterial proteins to the extracellular matrix of the biofilm of dental plaque, including van der Waals forces, hydrogen bonding, calcium linkages, and hydrophobic interactions (Sterzenbach et al. 2020). At present, we can merely speculate on the presence of amyloid fibrils or oligomeric aggregates as a possible mechanism of protein preservation, as before now, it has only been studied to a limited degree (Petzold et al. 2020). Although beyond the scope of this thesis, we propose that experimental studies on AF formation may be a good focus for future research. While the long-term preservation of proteins is likely to occur through multiple mechanisms, given that proteins are an enormously diverse group of biomolecules, the ability of some proteins to aggregate and form highly robust structures may offer a potential explanation for the overrepresentation of specific dietary proteins, such as BLG, compared to more abundant proteins from the same food sources such as $AS_1$-CN. Specifically, it may also help to explain why certain protein regions are more proteotypic and have substantially higher recoverability, as empirically seen for the large number of peptide spectral matches (PSMs) typically recovered for specific BLG peptides, such as the Bovid peptide $_{141}$TPEVDDEALEKFD$_{153}$ compared to others (Jeong et al. 2018) These and other properties and characteristics have been previously proposed to play a role in the long-term preservation of proteins and this will be further explored in the following section.

## 8.2 Patterns in protein preservation

It has been well-established that biomineralization of proteins can lead to exceptional long-term preservation in the archaeological record as mineralization likely plays an important protective role (Demarchi 2020). However, the exact mechanisms involved in the biomineralization of bone and the role of each protein are still unknown (Beniash 2011), and protein preservation and recovery of proteins within archaeological samples can still be highly unpredictable, even from mineralized material. Although the exact mechanisms involved in long-term protein preservation are still unknown, several protein characteristics have been proposed to aid in the preservation of archaeological proteins, including structure, calcium-binding ability (Collins et al. 2002; Warinner et al. 2014a; Demarchi et al. 2016; Hendy et al. 2018b), or specific PTMs (Ozcan et al. 2014).

Paleoproteomic studies often recover fibrous, insoluble proteins such as keratin, collagen, and elastins. These proteins are found in various tissues, including bone, connective tissue, wool, hair, and skin. Mineralized fibrous proteins have been demonstrated to be exceptionally well-preserved, although many of these proteins are highly conserved and are challenging to distinguish taxonomically. Additionally, collagens and keratins are common contaminants in laboratory blanks. In general, fibrous proteins are extremely strong structural proteins, and collagen's triple helical structure is exceptionally stable. Many collagen types form robust macromolecular fibrillar structures (Shoulders and Raines 2009). These structural properties may explain the frequency of collagen protein recovery from dental calculus in previous paleoproteomic studies, and they were frequently recovered in our samples as well. During periodontal disease, bone loss occurs, and collagen is shed into the oral environments and can

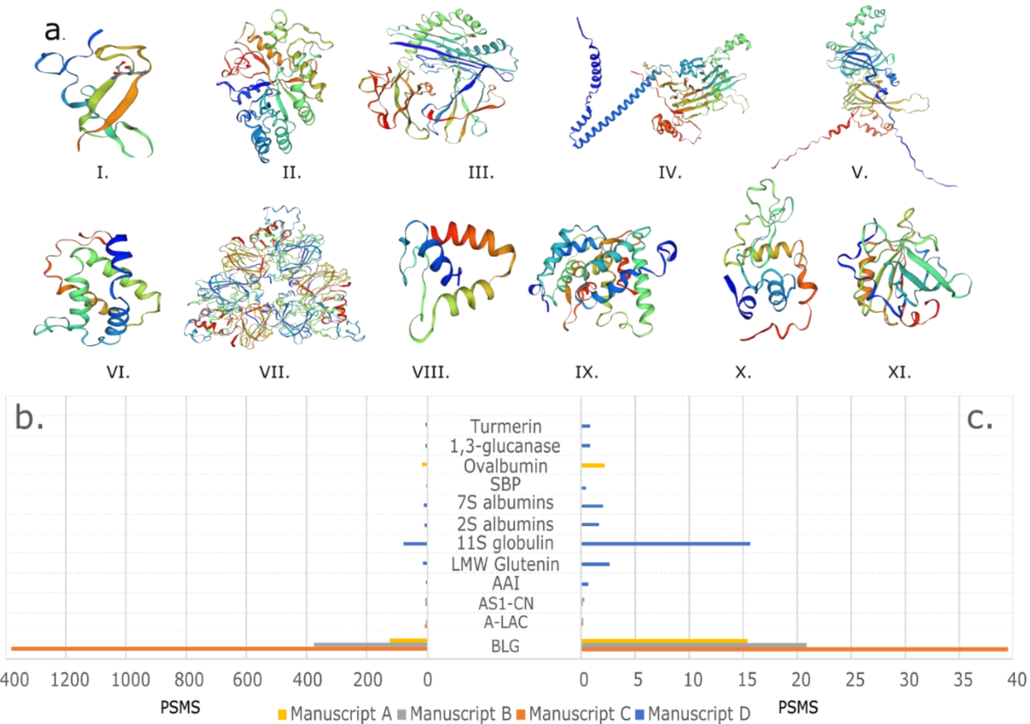



**Figure 4. 3D structures for and peptide spectral matches identified dietary proteins.** (a) 3D structures were downloaded from the SWISS-MODEL server and repository (Waterhouse et al. 2018; Bienert et al. 2016). (I) *Curcuma*, turmerin (II) *Musa*, β-1,3 glucanase (III) *Gallus*, Ovalbumin (IV) *Glycine*, sucrose-binding protein (SBP) (V) *Glycine*, 7S globulin (VI) *Sesamum*, 2S albumin (VII) *Sesamum*, 11S globulin (VIII) *Triticum/Aegilops*, low molecular weight (LMW) glutenin (IX) *Triticum/Aegilops*, α-amylase/trypsin inhibitor (AAI). (X) Bovine, α-lactalbumin (A-LAC) (XI) *Equus*, β-lactoglobulin (BLG) (b) raw count of PSMs for dietary proteins (c) PSMs normalized by sample size per study

easily become incorporated into the dental calculus. Additionally, during the removal of dental calculus from the surface of archaeological teeth, overaggressive scraping can also lead to an abundance of collagen from the tooth. While not all collagen is considered a contaminant, and is often an endogenous oral protein, these proteins were of limited value in our dietary reconstruction studies and were not examined further. In contrast to the recovery of fibrous insoluble proteins, the recovery of taxonomically distinct dietary protein from the archaeological record has been uncommon. In dental calculus studies, these identifications have consisted primarily of milk proteins, with BLG, in particular, dominating the finding of most studies, although this may be due to the fact that the studied populations engaged in pastoral-based subsistence strategies (Yang et al. 2014; Jeong et al. 2018; Charlton et al. 2019; Wilkin et al. 2020; Bleasdale et al. 2021).

To date, the limited amounts of non-collagenous paleoproteomic data available has made it difficult to formulate hypotheses about factors contributing to dietary protein preservation. Prior to this thesis research, BLG was the only dietary protein that had been recovered at high levels within individual calculus samples, and it has been recovered frequently and consistently in known dairy-consuming populations. While the preservation of proteins from archaeological bone has been attributed to their structural properties, native BLG, a soluble globular protein, does not share these robust structural characteristics. While BLG proteins from archaeological dental calculus share a calcium-binding capability with proteins such as

collagen, it shares few characteristics beyond the fact that they both become embedded within a calcium phosphate mineral matrix. Due to this, it has been proposed that the calcified mineral matrix itself plays a protective role in preserving BLG. In our datasets, we find high levels of BLG peptides in known pastoralist populations and the absence of milk proteins in populations that did not engage in dairying. In many cases, the protein is one of the highest-scoring within the calculus samples, and calcium-binding itself cannot explain its prevalence relative to other soluble globular milk proteins or other calcium-binding fibrous proteins that dominate most paleoproteomic studies of dental calculus.

In the four manuscripts comprising this thesis, we identified 20 dietary-specific proteins, including 13 non-dairy proteins (see Figure 4a). While BLG was still by far the most frequently recovered dietary protein in our dataset, this was only true for pastoralist populations that relied heavily on their livestock herds for subsistence (Manuscript B and D), seen in Figure 4b and 4c. The shared characteristics among the dietary proteins in our dataset are shown in Table 3. Regarding structure, all proteins were globular proteins or intrinsically disordered proteins (IDPs). Although caseins are classified as IDPs because they lack tertiary structures in their native state, they also form colloid structures that resemble denatured globular proteins. All identified dairy proteins are potential allergens with low molecular monomeric weights that are resistant to degradation through proteases, changes in pH, or heat. Each protein contained β-sheet secondary structures in its native form or could form them in specific conditions. Each protein contained calcium-binding, immunoglobulin E (IgE)-binding, or disordered domains. All the plant proteins were involved in storage or defense, as were many of the animal proteins, and many proteins have known antimicrobial properties that could also aid in preservation. The relationship between these characteristics and protein aggregation and AF formation is visualized in Figure 5 and will be discussed throughout the following sections.

Next, we evaluated previously reported archaeological proteins to determine if they also shared these characteristics. Although a large-scale evaluation of archaeological proteomic datasets is beyond the scope of this thesis, a few select studies were examined from different archaeological contexts, including proteomic studies of ancient eggshells, mollusk shells, pottery, and dental calculus, discussed in section 8.2.5. Although some of the reported archaeological proteins were identified as matches to predicted proteins automatically translated from genetic and genomic sequencing projects and therefore without available detailed secondary structural information, some structural information could still be inferred based on similarity to homologous proteins. Many of the evaluated proteins share properties with proteins from our dataset that are associated with AFs, and we propose that this may be a contributing factor in archaeological protein preservation

### *8.2.1. Amyloid fibrils*

*8.2.1.a. Functional amyloid fibrils are biologically important, widespread, and diverse*
These highly stable structures contain a cross-β scaffold pattern formed from intermolecular β-sheets, stabilized by an abundance of hydrogen bonds, and they are associated with resistance to protease enzymes and ionic detergents (Antonets et al. 2020). The ability to form amyloid structures was once believed to be rare, occurring only within a few proteins associated with amyloid-related neurodegenerative diseases and were thought to possibly Occur as a result of specific structural properties (Cao and Mezzenga 2019). More recent

**Table 3. Recovered dietary proteins and their shared characteristics**

| | Globular protein | Defense proteins | Storage proteins | Calcium-binding sites | Glycoprotein | Phosphoprotein | Disordered proteins or regions | β - sheets/ strands | β-Barrel / Cupin domains | IgE binding sites | Potential allergen | Low molecular weight (<70KDa) | Previously demonstrated amyloid fibrils formation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LMW-GS | * | | X | | X | | X | X | | X | X | X | X |
| AAI | X | X | | | | | | X | | X | X | X | |
| B-GLU | X | X | | | | | | X | | X | X | X | |
| Ovalbumin | X | X | X | X | X | X | | X | | X | X | X | X |
| Turmerin | X | X | | | | | | X | | X | X | X | |
| 11S globulins | X | X | X | | | | X | X | X | X | X | X | X |
| 2S albumins | X | X | X | | | | | X | | X | X | X | |
| 7S globulins | X | X | X | X | X | | X | X | X | X | X | X | X |
| SBP | X | X | X | | | | X | X | X | X | X | X | X |
| a-LACs | X | X | | X | X | | | X | | X | X | X | X |
| AS$_1$-CNs | * | X | X | X | | X | X | X | | X | X | X | X |
| BLGs | X | X | | X | | | | X | X | X | X | X | X |

*Intrinsically disordered proteins

research has demonstrated that amyloid fibril formation is a potential property of most proteins, and 98.7% of proteins contain at least one self-complementary short sequence capable of forming amyloid fibrils. However, most of these sequences are buried in the native fold of proteins or exist in structures that prevent the formation of the β-strand structures andinterdigitation needed for AF structures (Goldschmidt et al. 2010).

Non-pathological or functional AF formations appear to be an evolutionary adaptation shared across many kingdoms of life, with roles in nutrient storage and structural adhesion. These AFs play many functional roles, including human melanin production, human hormone accumulation and storage, nutrient storage and defense in fish eggs and insect eggs, the adhesive abilities of fungi, microbial biofilm formation, and the long-term storage of plant seed proteins (Antonets et al. 2020). The ability to self-assemble is a property that amyloidogenic proteins share with functional fibrous proteins such as collagen and keratin (Santos and Ventura 2021). The AF structures allow globular proteins to resemble these fibrous proteins more closely, granting them similar structural properties that increase their robustness. Biomaterial studies have even produced layered structures that included hydroxyapatite that could mimic bone, with AFs replacing collagen (Ke et al. 2020). Structural fibrous proteins within bone have been demonstrated to survive long-term in the archaeological record, and it has been theorized that their insoluble and robust fibril structures likely increased their chances of survival. If soluble globular proteins can also form similar and even more robust structures, it is possible that they can persist in this form over archaeological periods, particularly if those structures were trapped within a subsequently biomineralized matrix such as dental calculus.

*8.2.1.b. Amyloid fibrils are extremely stable and highly resistant to degradation*

Although morphology is variable, all amyloid fibrils share a cross-β sheet motif and structural repetitiveness where β-sheets lay parallel to the fibril axis and β-strands within individual sheets are arranged perpendicular to the central fibril axis (Cao and Mezzenga 2019). A dual β-sheets structure, also called a "steric zipper," is stabilized with additional forces including van der Waals, hydrogen bonds, and hydrophobic effects, resulting in the most tightly packed protein complexes known (Nelson et al. 2005; Sawaya et al. 2007) and the tensile strength of AFs have been compared to steel (Smith et al. 2006). These structural properties confer extraordinary stability to amyloid fibrils, increasing their resistance to degradation and survivability within human tissues, which could also increase their long-term survival. The presence of amyloid fibrils can be confirmed through the observation of the characteristic cross-β sheet x-ray diffraction patterns, as well as an observation of fibrillar morphology, AF-specific staining with Congo red (CR) or Thioflavin T (TnT), in conjunction with the observance of a characteristic apple-green birefringence under cross-polarized light microscopy (Ke et al. 2020).

Amylogenic, or core, regions of proteins promote β-sheet structure formation during fibrillation (Jansens et al. 2019a). These regions are typically short, highly hydrophobic sequences with low net charges (De Baets et al. 2014). Specific food protein fibril formation is affected by various factors, including amino acid sequence, charge, hydrophobicity, pH, time, temperature, ionic strength, and the presence of alcohols or enzymes (Jansens et al. 2019a; Lambrecht et al. 2019).

Depending on the amino acid sequences, the highly repetitive structure of AFs combined with charged and hydrophobic surface properties make them "sticky," which allows them to bind to various small molecules and macromolecules or polymers (Jacob et al. 2016). The extraordinary binding and stability properties of AFs contribute to the pathology of human diseases such as Alzheimer's, in which insoluble amyloid plaques accumulate between neurons, disrupting their function. However, functional AFs have been discovered to play a role in adhesion for organisms across kingdoms, and a study by Jacob et al. demonstrated that the cell-adhesive properties of AFs were intrinsic to the structure and not a result of the underlying sequence (Jacob et al. 2016). Many microorganisms, including bacterial and fungi species, use the adhesive properties of amyloids for surface attachment and colonization. Functional AFs are also closely associated with bacterial cell walls. Oli et al. demonstrated the importance of amyloid-forming proteins in the biofilm development of *Streptococcus mutans,* a prominent organism in the human oral cavity (Oli et al. 2012). Although several *S. mutans* proteins can form amyloid structures that are important structural components of biofilm, it remains unclear exactly how AFs affect the plaque structure, or which other amyloid fibrils are present within the plaque matrix (Jakubovics et al. 2021). Therefore, it may be that these structures play a role in the ability of plaque to entrap dietary proteins or the ability of dietary amyloid aggregates to attach to dental plaque on an individual's tooth surface before subsequent biomineralization.

*8.2.1.c. Amyloid fibril polymorphism*

Generally, amyloid fibrils exhibit a high degree of heterogeneity depending on experimental conditions. For example, BLG has been shown to form various polymorphic amyloid fibril structures depending on conditions such as ionic strength, pH, the degree of unfolding, and protein concentration (van den Akker et al. 2014). Full-length protein monomers and large peptides have been recovered from BLG amyloid fibrils heated to 80°C, and large amyloid-like fibrils formed from a combination of BLG, and other proteins have also been observed

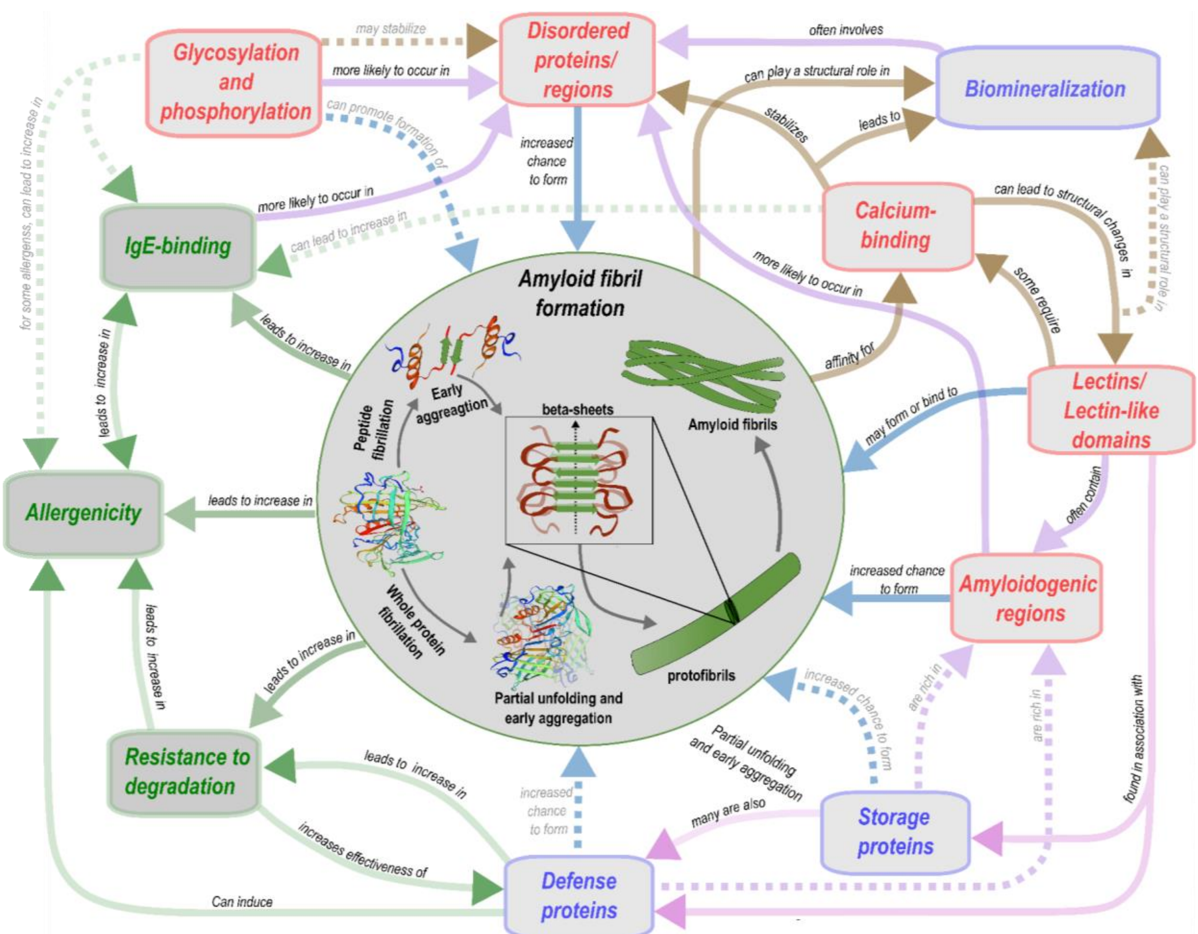


**Figure 5. Relationship between amyloid fibril formation and other protein characteristics**. Boxes describe the relationship between aggregation or AF formation and allergenicity (green), structural protein characteristics (red), functional protein characteristics (blue).

(Loveday et al. 2017). Mixtures of fibrils and non-fibrillar aggregates can be found within the same conditions. Mixtures of peptide fragments of varying lengths are expected in low pH conditions, but peptide fragments typically correspond to β-sheet rich regions BLG CITE. In addition to the abundance of polymorphic forms, many amyloid proteins have cross-polymerization properties in which they can form amyloid fibrils with different proteins, which is more likely to occur when similar peptides are involved. Others have been shown to form cross-species fibrils, including those involved in cross-species prion diseases, which could increase the likelihood of additional proteins being incorporated and possibly preserved within the fibril structures.

*8.2.1.d. Amyloid fibrils from dietary proteins*

Functional amyloid fibrils have been discovered in association with food proteins from dietary items such as milk, eggs, and seeds, which were also identified in our studies. They form under various *in vitro* fibrillation conditions, usually involving temperature, pH, salt, and protein concentration (Cao and Mezzenga 2019). These fibrils share the same cross-β pattern structure as pathological amyloid fibrils but are unrelated to any amyloidosis. However, studies on food protein fibrillation remain rare, and the field is not fully established. Plant proteins, in particular, have only been minimally investigated to date (Cao and Mezzenga 2019; Antonets et al. 2020).

The formation of amyloid or amyloid-like fibrils typically requires at least a partial unfolding of the native protein state to expose the hydrophobic core of soluble globular proteins. This is typically achieved through heating, low pH, or dehydration, methods common in food processing. However, prior to 2020 (Monge-Morera et al. 2020), the only amyloid fibrils unequivocally shown in prepared foods were in foie gras from duck and goose (Solomon et al. 2007). Most insights about dietary amyloid fibrils have resulted from *in vitro* studies of food proteins in purified forms, often to elucidate optimal fibrillation conditions of specific proteins for applications in food technologies. However, non-optimal conditions do not preclude fibril formation. To date, no studies have investigated amyloid formation in real-world complex food systems in which a taxonomic mixture of proteins may be present or how food processing methods may affect, prevent, or induce these formations. However, it is known that some specific proteins and food constituents can affect other proteins, either through reduction or enhancement of fibrillation. For example, *in vitro* studies of milk proteins show that some caseins may help prevent the fibrillation of whey proteins in some conditions (Lambrecht et al. 2019). Although recent studies have demonstrated the ability of some dietary proteins to form fibrils during common food-processing steps (Lambrecht et al. 2019, 2021; Monge-Morera et al. 2020, 2021a, b), more research is required to establish the presence of amyloids in other common food products.

*8.2.1.e. Amyloid fibrils of animal-derived dietary proteins*

Recent studies by Monge-Morera et al. (Monge-Morera et al. 2020, 2021b) have verified that fibrils are formed during both drying and boiling of hen (*Gallus gallus domesticus*) eggs. The authors found that heating egg white proteins (primarily composed of ovalbumin, ovotransferrin, ovomucoid, ovomucin, and lysozyme) inducing the formation of intermolecular β-sheet structures, and 1-3% of egg white proteins were converted into amyloid fibers, and their results add to the growing body of research that suggests that amyloid fibrils are a common aspect of the human diet. Purified OVAL and LYZC proteins can also form amyloid aggregates at low pH (2.0) (Jansens et al. 2019a).

Amyloid formation of purified whey proteins such as BLG or A-LAC typically requires heating in low pH conditions for the optimal release of amyloidogenic peptides that can form β-sheet structures, although the presence of other denaturing chemistry can allow AF formation in slightly higher pH conditions (Lambrecht et al. 2019) In contrast, bovine serum albumin can form AFs at neutral pH with only mild heating. In general, whey proteins form AFs more easily than casein proteins (Lambrecht et al. 2019), which may partly explain why whey proteins are recovered more often than caseins in archaeological dental calculus. Caseins (CNs) are intrinsically disordered proteins lacking a consistent tertiary structure, but which form micelle aggregates. Although purified K-CN and $AS_2$-CN can form AFs even at physiological pH, purified $AS_1$-CN and B-CN do not readily form AFs until they are exposed to low pH (2.0) and high temperatures (90°C) conditions (Lambrecht et al. 2019). Additionally, various proteins, food constituents, and ions can impact fibril formation in milk proteins. BLG is large, and richer in β-sheets compared to other milk whey proteins, leading to more stabilizing intermolecular bonding. This fact may help partially explain the abundance of the protein in paleoproteomic studies in relation to other calcium-binding milk proteins. BLG is also the more heat-sensitive milk protein and will begin to unfold and aggregate at lower temperatures than A-LAC and caseins (Bu et al. 2013), which could increase the chances of amyloid formation.

*8.2.1.f Amyloid fibrils of plant-derived dietary proteins*

To date, very few investigations of plant amyloid biology have occurred. Antonets and Nizhnikov performed a large-scale bioinformatic screening of plant proteins from all

annotated plant proteomes, demonstrating the abundance of potentially amyloidogenic proteins in plant proteomes suggesting that AFs may play an important functional role (Antonets and Nizhnikov 2017). Potential AF-forming plant proteins include a high number of defense proteins and nutrient reservoir or seed storage proteins.

Many plant defense proteins have shown amyloid-like properties *in vitro* (Garvey et al. 2013; Gour et al. 2016; Berthelot et al. 2016). The stability of AFs may improve the survivability of defense proteins during pathogen interactions. Additionally, the Antonets and Nizhnikov study demonstrated the AF potential of nutrient reservoir proteins, particularly seed storage proteins, especially those enriched in Q and E, with conserved β-barrel domain Cupin-1, which is rich in amyloidogenic regions and present in most plant species (Antonets and Nizhnikov 2017; Antonets et al. 2020). In a follow-up study published in 2020, Antonets et al. demonstrated that Cupin superfamily seed storage protein, vicilin, could form more than one type of aggregate *in vitro* (Antonets et al. 2020). Every plant protein we identified plays a role in defense or storage, and several storage proteins containing Cupin domains were identified in our study described in manuscript D.

Antonets et al. were able to demonstrate that AFs formed both *in vitro* and *in vivo* in seed storage proteins, likely stabilizing these storage proteins and preventing degradation during dehydration/desiccation and seed dormancy (Antonets et al. 2020). Although there has been a distinct shortage of amyloid studies of cereal proteins, some studies have demonstrated that cereal proteins belonging to wheat gluten, maize, and rice exhibit AF forming properties in specific *in vitro* conditions (Jansens et al. 2019a). Recent studies by, Monge-Morera et al. and Lambrecht et al. demonstrated that wheat gluten proteins, which were also recovered in our studies, could be induced to form amyloid fibrils during food-processing relevant conditions including drying and heating (Monge-Morera et al. 2021a; Lambrecht et al. 2021).

### *8.2.2 Protein preservation and allergenicity*

Many known food allergens have been demonstrated to form AF in certain (and sometimes very specific) conditions, including β-parvalbumin from fish, whey and casein proteins from milk, ovalbumin, and lysozyme from egg, and many seed storage and defense proteins from plants. Each protein we recovered in our studies was a potential human allergen that contained IgE-binding sites required for IgE-mediated immune responses. Even though whole proteomes were available for some of our identified species, and others were extensively represented as they are important human dietary crops, only the allergenic proteins were identified, and an overrepresentation of human allergens in our database is, therefore, unlikely to be responsible for our results. Proteins capable of eliciting an IgE-mediated allergic reaction share many characteristics, although these characteristics are not unique to allergens and are also shared amongst non-allergenic proteins (Pekar et al. 2018). However, in general, these proteins have small molecular sizes (typically under 70 kDa) (Vickery et al. 2011), are stable against changes in pH, and are resistant to enzymatic degradation. They also often possess thermostability properties. Despite an abundance of research, the unique properties of allergenic proteins remain unknown (Pekar et al. 2018).

Phosphorylation or glycosylation can increase the allergenicity and IgE-binding activity of several allergen proteins (e.g., caseins, ovomucoids, serpins), including many identified in our studies (Costa et al. 2021). However, the opposite may be true for others (e.g., tropomyosin). Recently, amyloid aggregates of allergenic proteins have also been shown to enhance IgE recognition and binding (Sánchez et al. 2016; Castellanos et al. 2018), as shown in Figure 5. The recent discovery of *in vivo* AF structures of the highly allergenic vicilin seed

lectin proteins helps explain the protease-resistant properties associated with the plant's defense (Antonets et al. 2020). The extreme stability of the insoluble AF gives the proteins a heightened resistance to degradation, increasing their allergenicity. Research has demonstrated that IgE-binding is increased 1000-fold in β-parvalbumin AFs compared to the monomeric structure and that the IgE-binding epitopes overlap with the amyloid cores of the protein, which may explain the IgE binding properties of the amyloid structures (Pérez-Tavarez et al. 2021). This study also demonstrated that BLG and OVAL display the same pattern, both of which were identified in our datasets. Interestingly, the peptide most frequently recovered in paleoproteomic studies, $_{141}$TPEVDDEALEKFD$_{153}$, overlaps with the IgE-binding epitope and is flanked by two amyloid regions (see Fig. 6), while the second IgE-binding epitope region of BLG overlaps with another core-forming region. The other most frequently recovered BLG peptide in paleoproteomic studies (Jeong et al. 2018) is peptide $^{108}$ VLVLDTDYK $^{116}$, which falls within a region of predicted amyloid cores between both IgE-binding epitope regions

***8.2.3 Protein preservation and protein structure***

Some protein domains have been associated with amyloid fibril formation (see Figure 5), including those with low sequence complexity, disordered regions, β-sheet forming core regions, and carbohydrate calcium, and IgE-binding regions, and many of these were found in our dataset.

*8.2.3.a. Disordered regions*

Amyloid fibrils often occur in intrinsically disordered proteins or proteins that contain intrinsically disordered domains, as these regions are more susceptible to aggregation. Disordered proteins or regions that lack a defined three-dimensional structure are an important aspect of many processes involved in amyloid biology and possibly in the processes involved in the preservation of proteins. Conserved disordered regions often overlap in sequence with other protein domains (Pentony and Jones 2010). Due to disordered regions' polar charged amino acid stretches, these regions may generally increase binding rates since these sequences are not hydrophobic enough for folding. Additionally, disordered regions can allow proteins to interact with domains of other proteins (van der Lee et al. 2014). IgE-binding sites often co-occur within or adjacent to disordered regions of proteins (Xue et al. 2011). Phosphorylation and glycosylation are common post- translational modifications in disordered regions of animal-derived allergens (Costa et al. 2021; Gattinger et al. 2021), and they may have a stabilizing effect on the structure of the proteins (Grzybowska 2018). A number of glycoproteins also play a role in biomineralization (Evans 2020). The physicochemical processes that allow disordered regions to form ordered, β-sheet rich fibrils are not fully understood but may involve binding ligands or metal ions (Lermyte 2020), and structural disorder is also common in proteins that play a role in the biomineralization of hard tissue. Disordered proteins, or proteins with disordered regions recovered in our studies, include glutenin, 11S and 7S globulins, sucrose-binding protein, and $AS_1$-CN.

Amyloidogenic regions are often associated with other protein domains, including IgE-binding, calcium-binding, and carbohydrate-binding regions. For example, the conserved seed storage Cupin-1 β-barrel domain contains an abundance of amyloidogenic regions in most plant species

*8.2.3.b. Amyloidogenic regions*

Amyloidogenic regions promote the β-sheet structure formation required for fibrillation [138],

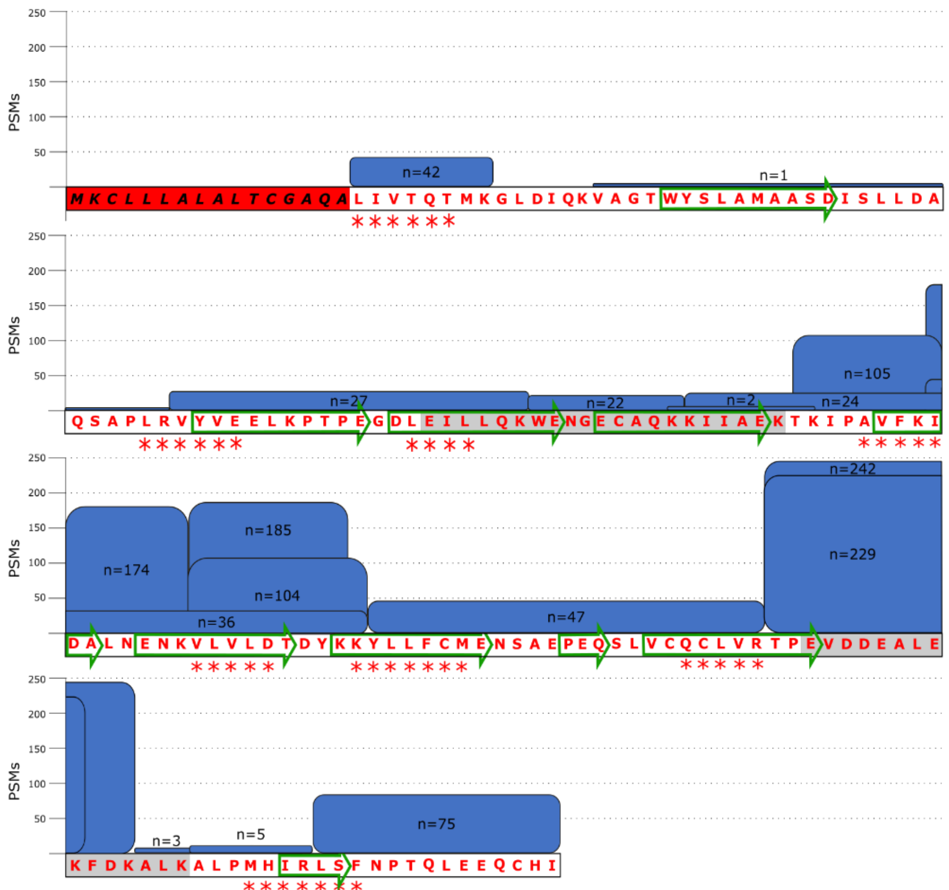



**Figure 6. Recovered BLG peptides in relation to IgE binding sites and predicted amyloid aggregation regions.** Blue boxes represent BLG peptides recovered in Manuscript C with the number of PSMs identified in the entire dataset. The N-terminal region shaded red is the signal peptide. Areas shaded in grey correspond to IgE-binding epitopes of BLG (Jansens et al. 2019b; Pérez-Tavarez et al. 2021). Arrows represent β-sheet secondary structures and asterisk (*) corresponds to aggregation prone segments, predicted using AMYLPRED2 software (Tsolis et al. 2013). It should be noted that these regions are not discrete units and the actual protein sequences included in any given amyloid fibril can be expected to include regions outside of the sequences responsible for the aggregation (denoted with an asterisk *), or the regions between these sequences in the case of whole protein, or large sequences aggregation and fibrillation.

and they often occur in or near disordered regions of proteins. Many proteins are involved in defense, and their homologs in marine species such as oysters have been demonstrated to play a role in amyloid fibrillation (Antonets et al. 2020; Naganuma et al. 2014). The β-barrel structures, formed from β-sheets, are particularly rigid and stable. In fact, many plant lectin proteins form the same compact β-barrel structure. Interestingly, proteins belonging to the lipocalin protein family, including β-lactoglobulin, also contain a similarly conserved antiparallel β-barrel structure (Crowther et al. 2016). This may explain the frequency of recovered peptides that correspond to β-sheet rich regions of BLG and 11S globulins peptides that were recovered in our studies, constituting the most frequently recovered peptides in our dataset (see Figure 4b and 4c). If this β-barrel structure does help increase stability and the likelihood of preservation, it may also help explain the differentially high recovery of BLG peptides relative to other milk peptides.

*8.2.3.c. Lectin and Lectin-like proteins and domains*

Carbohydrate-binding proteins can play an important structural role in AF formation by providing scaffolding in various contexts, including in bacterial biofilms (Oli et al. 2012), oyster shells (Naganuma et al. 2014), plant seeds (Antonets et al. 2020; Santos and Ventura 2021), and in diseased tissue of amyloidosis patients (Iannuzzi et al. 2015). Several archaeological proteomic studies have reported lectins or proteins with lectin-like properties (Demarchi et al. 2016; Presslee et al. 2017; Sakalauskaite et al. 2019). The seed storage proteins in our dataset have lectin-like properties due to the conserved Cupin domain, which has an affinity for carbohydrate-binding. It is primarily present in 11S and 7S globulin seed storage proteins, which are major allergens common in human diets. The QN-rich domain belonging to plant zein, gliadin, vicilin, and glutenin proteins are also potentially amyloidogenic regions. Many lectin proteins in plants and animals exhibit defensive properties against microorganisms (Dias et al. 2015) in addition to their nutrient storage roles. Although the physiological processes involved in seed protein preservation have been poorly understood, Antonets et al. demonstrated the *in vivo* involvement of AFs in storage protein accumulation within plant seeds. The presence of these highly robust structures within seed proteins *in vivo* may help explain the high levels in which they were recovered in our dataset, as they would not require the same extreme food-processing conditions required by other proteins to form AFs.

Although the role of lectins in human allergies is still poorly understood (Barre et al. 2020), several proteins containing lectin or lectin-like proteins have been identified as potential allergens, including the Cupin domain-containing proteins identified in this thesis. While calcium does not participate directly in the carbohydrate-binding of lectin proteins, it is required by some, including those containing C-type lectin (CTL) domains or the homologous C-type lectin-like domains, referred to as CTL throughout the remainder of this thesis. The CTL domains have four calcium-binding regions, although a variable number is occupied depending on structure (Zelensky and Gready 2005). Calcium-binding in these domains can lead to structural changes in the protein that have been associated with scaffolding for biomineralization and host defense against pathogens. In some cases, the conformational changes increase interaction with IgE-binding sites (Engeroff and Vogel 2021).

*8.2.3.d. Calcium-binding*

Many previously identified archaeological proteins have calcium-binding capabilities, leading to the hypothesis that mineral binding itself may play a protective role in protein preservation (Warinner et al. 2014a; Demarchi et al. 2016; Hendy et al. 2018a; Demarchi 2020). Calcium-binding proteins have a higher tendency to form the extended β-strand structures found in amyloid fibrils that increase their aggregation capacity, giving the AF structures an affinity for calcium ions and subsequent calcification. Some non-collagenous proteins in dentin and bone, such as DMPI and phosphophoryn, can form amyloid-like fibrils when bound to calcium, likely due to the β-sheet conformation adopted by the proteins in the presence of the calcium ions, and these aggregates of non-collagenous proteins may be vital for the regulation of mineral formation and organization in bone tissues (Grzybowska 2018).

Calcium-binding can also play a role in the allergenicity of some proteins. For example, in the case of β-parvalbumin, calcium-binding leads to structural changes that leave its IgE-binding regions available, which conserves its allergenic potential, and even more so in subsequent amyloid fibrils (Sánchez et al. 2016).

### *8.2.4 Protein preservation and protein function*

Functional amyloid fibrils play a key role in a wide variety of biological processes, including functions that may enhance their chances of long-term survival, such as biomineralization, defense, and storage, as shown in Figure 5.

*8.2.4.a. Biomineralization*

Calcium-binding domains within proteins promote disorder (Grzybowska 2018), and intrinsically disordered proteins (or proteins containing disordered domains) are thought to be important in initiating biomineralization (Boskey and Villarreal-Ramirez 2016). The frequency of phosphorylation and glycosylation in disordered regions can stabilize the structure when coupled with the surface adherence needed to initiate calcification. This may be due to the lower free energy required for disordered proteins due to a change in entropy while the protein structure transitions from disordered to ordered (Grzybowska 2018).

Most biomineralization involves calcium phosphate, calcium carbonate, or hydroxyapatite, and they are involved in the mineralization of mollusk shells, eggshells, bones, teeth, and dental calculus. Castanellos et al. demonstrated that amyloid aggregates of β-parvalbumin from Atlantic cod muscle could influence calcium carbonate precipitation *in vitro* (Castellanos et al. 2018*).* Interestingly, they found that amyloid aggregates alone were able to alter calcite morphology. The ability of amyloids to modulate calcium carbonate crystallization is due to their polymorphic structures and varies with cellular conditions and covalent structures of the protein. Due to their strong adhesion and bond stability, the amyloid structures of self-assembled peptides can act as templates that induce hydroxyapatite formation and layering (Wang et al. 2021).

The β-parvalbumin protein shares many characteristics with our identified proteins; it is a globular allergen that binds both calcium and IgE. Naganuma et al. demonstrated that some lectin proteins (PPL2s) and their AF aggregates contribute to the biomineralization process in pearl oyster mollusk shells as matrix proteins. These lectin proteins are homologous to a specific lectin protein in plants, one of the types of proteins predicted to form AF aggregates more readily (Naganuma et al. 2014; Antonets and Nizhnikov 2017). In contrast to studies that demonstrate AF aggregation proceeds biomineralization, research by Pal et al. demonstrated that biomineralization of alkali halides (KCl and NaCl) could induce AF formation of the protein ovalbumin via changes in salt concentration *in vitro* (Pal et al. 2011). Increased salt concentrations can lead to increased β-sheet formation as the protein unfolds, leading to fibrillation due to the biomineralization of the alkali halides.

*8.2.4.b. Storage and Defense*

All dietary plant proteins identified in this thesis had roles in either storage, defense, or both, and these functional classes are overrepresented in predictive studies of potential amyloidogenic sequences (Antonets and Nizhnikov 2017). Our dataset includes seed storage proteins from the prolamin and Cupin superfamilies, including 11s and 7S globulin seed storage proteins, sucrose-binding protein, 2S albumin, glutenin, and α-amylase/trypsin inhibitor protein (AAI). Milk caseins and the serpin glycoprotein ovalbumin were also recovered. They are considered storage proteins as they provide nutrients for growing embryos.

Defensive proteins in our dataset include two glycosyl hydrolase proteins, A-LAC, and β-1,3 glucanase. These proteins hydrolyze the glycosidic bond of carbohydrates and play a role in plant defense against pathogens by hydrolyzing bonds within prokaryotic cell walls.

Turmerin and AAI proteins are serine protease inhibitors that protect the plant from the digestive proteases of insects. Identified milk proteins include BLG, A-LAC, and $AS_1$-CN proteins, which display antimicrobial properties (Chaneton et al. 2011; Dziuba and Dziuba 2014; Hou et al. 2018).

### *8.2.5 Shared characteristics from other paleoproteomic studies*

*8.2.5.a. Amelogenin*

Archaeological studies have demonstrated the exceptional preservation potential of proteins within dental enamel (Nielsen-Marsh et al. 2009). This mineralized tissue is exceptionally strong and does not regenerate during life. Although there are few proteins within this tissue, amelogenin is abundant and valuable for archaeological studies. Amelogenin is encoded by the X and Y chromosomes, with different variants, which allows some insight into biological sex. This intrinsically disordered protein binds both phosphate and calcium ions (Stewart et al. 2017; Parker et al. 2019b) and has been demonstrated to be exceptionally well preserved (Capellini et al. 2019; Welker et al. 2020) despite its native structural properties.

Additionally, due to its long-term preservation potential, amelogenin has been utilized in paleoproteomic studies of human evolution (Welker et al. 2020). As noted by Welker et al., amelogenin is proteolytically cleaved during enamel formation by matrix metalloproteinase-20 (MMP20) and expected peptide cleavage products were recovered in their study. However, it should be noted that these *in vivo* cleavage products have been demonstrated to form AF structures subsequently. Carneiro et al. 2016 demonstrated that full-length amelogenin and its MMP20 cleavage products aggregate into amyloid-like structures both *in vivo* and *in vitro*. They identified the specific structure within amelogenin that formed into β-sheets and proposed that the amyloid structures play an essential structural role in guiding the apatite mineralization of tooth enamel (Carneiro et al. 2016). Zhang et al. demonstrated the core sequence $_{18}$EVLTPLKWYQSI$_{29}$ as the likely primary segment that contributes to the formation of the secondary structure required for AF formation (Zhang et al. 2020). This region was also recovered by Welker et al, and utilized sequence difference to distinguish between early hominins.

In the Supplementary Information supplied by Welker et al., two archaeological protocols were compared for their efficiency in recovering enamel proteins, including enamelin, both amelogenin variants, ameloblastin, and amelotin. The first method uses a dilute 4% HCl pretreatment, while the second method uses a much more concentrated 1.2M HCl pre-treatment. The second method was more effective in recovering more sequence coverage of these proteins, but the differences were much more apparent in the amelogenin proteins, the only proteins that are known to form AFs in enamel. Due to the protective properties of the AF structure, acidic pre-treatment is required for depolymerization; therefore, the differences in sequence coverage could plausibly be a result of more successful depolymerization using the second method listed.

*8.2.5.b. Avian egg proteins*

The oldest identified avian eggshell protein, ostrich struthiocalcin, dated to ~3.9 million years ago (Demarchi et al. 2016), recovered from fragments of archaeological eggshells. This C-type (calcium-binding) lectin protein possesses calcium-binding disordered regions that adopt a stable conformation when bound to a calcium source, a property shared with many of our protein identifications. Researchers have also demonstrated the role of modern struthiocalcin proteins in the initiation of calcification, and they make up the major components of the

eggshell matrix (Mann and Siedler 2004; Ruiz-Arellano et al. 2015).

Although amyloid fibrillation is likely to be an important mechanism in the biomineralization of mollusk shells and microbial biofilms, it has not been shown to be involved in the calcium-binding of avian eggshells. However, AFs have been demonstrated to play a protective role in the storage abilities of eggshells of silkworms, protecting the proteins from environmental factors even in the absence of calcified eggshells (Iconomidou et al. 2000), and AFs may play an important role in the preservation of storage proteins.

Other struthiocalcin-like lectin proteins involved in the biomineralization of avian eggshells and egg white proteins have also been reported from archaeological eggshells. For example, the homologous proteins ansocalcin and ovocleidin-17, found in goose and chicken, respectively, have been recovered (Presslee et al. 2017; Demarchi et al. 2019, 2020), and both proteins share significant homology with another C-type lectin (CTL) found in snake venom (Lakshminarayanan and Kini 2002). The AF forming properties of CTL proteins in snake venom play a vital role in the aggregation properties of the protein through their extensive β-sheets structures (Aranda-Souza et al. 2019). The CTL ovocleidin-17 and ovocleidin-like proteins have antimicrobial properties, and they likely play a protective role for the egg's embryo (Wellman-Labadie et al. 2008).

Egg white proteins such as ovomucoid, ovotransferrin, ovalbumin, and ovostatin have also been reported in archaeological eggshells (Presslee et al. 2017; Demarchi et al. 2019, 2020). These proteins are known to adopt AF structures under certain conditions such as heating, drying, or low pH. Both ovomucoid and ovomucin are glycoproteins, while ovomucoid, ovomucin, and ovalbumin are IgE-binding allergens that also play a role in bacterial defense (Colgrave 2017). Ovomucin forms fibrils that are believed to play a role in preventing bacterial movement within the egg white (Strixner and Kulozik 2011), while ovomucoid is a protease inhibitor protein (Colgrave 2017). Both proteins have β-sheet structures, and ovomucin contains disordered regions.

*8.2.5.c. Mollusk shell proteins*

Sakalauskaite et al. also identified several proteins from archaeological buttons made from mollusk shells (Sakalauskaite et al. 2019). Although data analysis was complicated due to insufficient database coverage of mollusk species, many proteins could be identified to the genus level. Several proteins involved in shell formation and biomineralization were reported, including Hic74, Hic 52, MSI60-related protein, silkmapin, and antifreeze protein. Although not as well characterized as dietary proteins, limited general structural information was available for some identified proteins. Hic 74, Hic 52, and MSI60 proteins have known disordered regions, while Hic74, MSI60, silkmapin, and antifreeze proteins have calcium-binding sites or structural motifs predicted to bind calcium (Liu et al. 2015, 2017b, a). The Hic 74 protein can form β-sheet structures that provide structural support for biomineralization (Liu et al. 2017a). Both MSI60 and silkmapin proteins display β-fold structures and are known to form fiber-like structures that likely play a role in initiating and growing biomineralized layers of mollusk shells (Liu et al. 2017a).

Sakalauskaite et al also identified an antifreeze protein from oyster (Sakalauskaite et al. 2019). Marine antifreeze glycoproteins share homology with calcium-binding plant lectins (Graether and Sykes 2004), and they also acquire stability when bound to calcium (Martinez et al. 2015). An important characteristic of antifreeze proteins in fish is the ice-induced amyloid fibril formation, once again demonstrating the ability of the AF structure to protect

proteins in extreme conditions, and Arai et al. have demonstrated that calcium-binding is a requirement for the ice-binding mechanism that induces AF formation in fish (Arai et al. 2019). Although the fish protein has a strictly α-helical native structure, it is converted into the β-oligomeric structures required to form amyloids. Although the exact mechanism remains unknown, it may be related to changes in pH, protein concentration, or specific interaction with ice (Graether and Sykes 2004; Dubé et al. 2016). Homologous insect antifreeze protein has the requisite β structures needed for AF formation (Graether and Sykes 2004).

*8.2.5.d. Plant proteins*

Few plant proteins have been reported to date in paleoproteomic studies, but a few examples include the same classes of seed storage proteins found within our datasets discussed above. These include oat (*Avena sativa*) 12S globulin (Jersie-Christensen et al. 2018), grape (*Vitis vinifera*) 2S albumin and 7S globulin (Cappellini et al. 2010), and pea (*Pisum sativum*) 11S globulin legumin proteins (Hendy et al. 2018a).

# 9. Conclusion

We combined our proteomic evidence with other analyses in three of the studies highlighted below, including bioarchaeology, microremain analysis, and genetics. Using a multidiscipline approach allowed for a deeper understanding of the diet of the individuals we examined. All data was contextualized with previous and ongoing archaeological evidence, including historical texts, archaeobotanical and zooarchaeological remains, and isotopic, genetic, and organic residue data. Through the application of paleoproteomics methods to archaeological dental calculus, we were able to:

1. Provide insight into the diets and living conditions of Ireland's poor and marginalized populations during the Great Famine of 1845-1852.
2. Demonstrate that indigenous individuals at the Bronze Age site of Xiaohe in the Tarim basin had adopted dairying technologies from neighboring pastoral groups and were already using these technologies by the earliest founding of the site despite maintaining genetic isolation from outside groups.
3. Identify the spread and diversification of dairy technologies in the Northern Caucasus, where an isotopically complex landscape and scant archaeological settlement data have hindered previous dietary studies. For the first time, our protein results established unequivocal evidence of dairy consumption in a region that is suspected to be a key area in the establishment and spread of high mobility dairy pastoralism that may have allowed subsequent migrations across Eurasia. Our results also revealed how humans responded and adapted to profound changes in environmental conditions of the steppe zone through time.
4. Recover a more diverse set of dietary proteins than previous dental calculus studies when studying an ancient urban environment rather than pastoralist groups that relied on such extreme herd-focused subsistence strategies. The tissue and genus-specific food proteins we recovered included those from fruit, spices, and seed proteins, some of which remain largely invisible in the archaeological record and would be unlikely to be detected through other methods. These proteins gave us insight into the interconnections between South Asia and the Mediterranean during the Bronze and Iron Ages.

Despite these successes, the recovery of taxonomically distinct non-dairy dietary proteins from calculus was rare. In cases where we were successful, these identifications were able to

contribute important insight into diet and trade that may have contributed to access to specific dietary species. Additionally, our dairy results continued to provide insight into the spread of dairy pastoralism contributed to the growing body of literature on dairying in the context of population genetics and movement and ancient environmental conditions.

Although we theorize that amyloid fibrils could form or incorporate into the dental calculus, we cannot confirm it through this thesis. While many aspects of amyloid biology, archaeological protein preservation, biomineralization, and dental calculus formation remain unknown, the possibility of amyloid fibrillation as a mechanism of protein preservation is a potential area of future research. Functional amyloid fibril research itself is still a newly emerging field, especially regarding animal and plant proteins related to human diet, and future research in these fields may help inform our hypotheses about dietary protein preservation.

Future AF studies of archaeological material can start this work by first confirming the presence of AFs in calculus. This might be achieved through the use of AF-specific staining in combination with polarizing light microscopy to confirm the characteristic apple-green birefringence associated with AFs. If AFs are present in archaeological dental calculus, it may be likely that we are under sampling them due to the greater difficulty of denaturing and digesting them. Although pre-amyloid structures and aggregates are often found alongside fully polymerized AF proteins, and these structures are still soluble in SDS, full AFs will likely require an acidic pre-treatment to depolymerize the structures before extraction. If present, calculus samples treated in increasingly acid concentrations of formic acid may show a differential protein extraction profile as proteins are progressively depolymerized. The issues associated with differential solubility of amyloid proteins and their retrieval from calcified dental calculus will require archaeological calculus-specific studies, although this presents the same challenges discussed in the previous section. Depending on the presence of any potential bacterial biofilm amyloid fibrils, the recovery of these proteins in high abundances might swamp out the signal of our dietary proteins. Empirical studies are needed to understand these factors better.

However, if AFs are demonstrated to play a role in preserving dietary proteins within dental calculus, targeted peptide studies may represent a path forward. In many of our proteins (if not all), our recovered dietary peptides corresponded to areas of the proteins that are predicted to be amyloid-forming regions through the predictive software AMYLPRED2 (Tsolis et al. 2013). The ability to predict aggregation-prone regions of natively globular and disordered proteins of interest, combined with known digestion cleavage sites, can help us choose appropriate peptides to targeted studies. While amyloid fibrillation would not be the only expected mechanism of preservation for highly diverse proteins, it could provide a productive starting point for further research.

While more studies are needed to confirm any pattern of dietary recovery found within this thesis, if the patterns hold up, we may be able to better predict the types of dietary proteins we can expect to be able to recover in future studies, which can help us make appropriate decisions about the allocation of scarce archaeological material. This may be especially important depending on the archaeological context and the aims of the research in question. If our patterns hold, future studies of dietary remains may yield improved results if they target storage and defense proteins, especially those with known allergenic properties. Seed storage proteins, in particular, are abundant and have now been demonstrated to survive in several contexts, including in calculus, pottery, and macrobotanical remains (Jersie-Christensen et al.

2018; Scott et al. 2021; Hendy et al. 2018a; Cappellini et al. 2010). Lectin proteins with antimicrobial properties may also have good preservation potential. Studies that focus on the recovery of peptides that correspond to predicted aggregation-prone regions of these proteins may be a valuable area of future study.

Although this thesis highlights a number of successful contributions to the field of paleodiet reconstruction via proteomic analysis of ancient dental calculus, many proteins we might expect to find based on the archaeological record were not recovered, indicating that our current method, while successful, is not complete and there is substantial room for improvement. Proteomic studies have enormous potential to add to our current understanding of paleodiet, and they offer specificity and statistical precision that other methods cannot. Further basic research into protein preservation and recovery will help this method reach its full potential.

# 10. References


Adler, C. J., K. Dobney, L. S. Weyrich, J. Kaidonis, J. Kaidonis, A. Walker, W. Haak, et al. 2013. “Sequencing Ancient Calcified Dental Plaque Shows Changes in Oral Microbiota with Dietary Shifts of the Neolithic and Industrial Revolutions.” *Nature* 45: 450–55.

Ahmed, Harby E., Ibrahim F. Tahoun, Ibrahim Elkholy, Adel B. Shehata, and Yassin Ziddan. 2017. “Identification of Natural Dyes in Rare Coptic Textile Using HPLC- DAD and Mass Spectroscopy in Museum of Faculty of Arts, Alexandria University, Egypt.” *Dyes and Pigments* 145: 486–92.

Akker, Corianne C. van den, Michael Schleeger, Mischa Bonn, and Gijsje H. Koenderink. 2014. “Structural Basis for the Polymorphism of β-Lactoglobulin Amyloid-like Fibrils.” In *Bio-Nanoimaging: Protein Misfolding and Aggregation*, edited by V. Uversky and Y. Lyubchenko, 333–43. Academic Press.

Albarella, Umberto, Mauro Rizzetto, Hannah Russ, Kim Vickers, and Sarah Viner-Daniels, eds. 2017. *The Oxford Handbook of Zooarchaeology*. Oxford University Press.

Ambrose, Stanley H. 1991. “Effects of Diet, Climate and Physiology on Nitrogen Isotope Abundances in Terrestrial Foodwebs.” *Journal of Archaeological Science* 18 (3): 293–317.

Ambrose, Stanley H., and Lynette Norr. 1993. “Experimental Evidence for the Relationship of the Carbon Isotope Ratios of Whole Diet and Dietary Protein to Those of Bone Collagen and Carbonate.” In *Prehistoric Human Bone: Archaeology at the Molecular Level*, edited by Joseph B. Lambert and Gisela Grupe, 1–37. Berlin, Heidelberg: Springer Berlin Heidelberg.

Antonets, Kirill S., and Anton A. Nizhnikov. 2017. “Predicting Amyloidogenic Proteins in the Proteomes of Plants.” *International Journal of Molecular Sciences* 18 (10). https://doi.org/10.3390/ijms18102155.

Antonets, K. S., M. V. Belousov, and A. I. Sulatskaya. 2020. “Accumulation of Storage Proteins in Plant Seeds Is Mediated by Amyloid Formation.” *PLoS Biology* 18 (7): e30000564.

Arai, Tatsuya, Yoshiyuki Nishimiya, Yasushi Ohyama, Hidemasa Kondo, and Sakae Tsuda. 2019. “Calcium-Binding Generates the Semi-Clathrate Waters on a Type II Antifreeze Protein to Adsorb onto an Ice Crystal Surface.” *Biomolecules* 9 (5). https://doi.org/10.3390/biom9050162.

Aranda-Souza, Mary Ângela, Vírginia Maria Barros de Lorena, Maria Tereza Dos Santos Correia, Antonio Pereira-Neves, and Regina Celia Bressan Queiroz de Figueiredo. 2019. “A C-Type Lectin from Bothrops Leucurus Snake Venom Forms Amyloid-like Aggregates in RPMI Medium and Are Efficiently Phagocytosed by Peritoneal Macrophages.” *Toxicon: Official Journal of the International Society on Toxinology* 157: 93–100.

Austin, J. J., A. J. Ross, A. B. Smith, R. A. Fortey, and R. H. Thomas. 1997. “Problems of Reproducibility--Does Geologically Ancient DNA Survive in Amber-Preserved Insects?” *Proceedings. Biological Sciences / The Royal Society* 264 (1381): 467–74.

Barker, A., J. Dombrosky, D. Chaput, and B. Venables. 2015. "Validation of a Non-Targeted LC-MS Approach for Identifying Ancient Proteins: Method Development on Bone to Improve Artifact Residue Analysis." *Ethnobiology*.

Barker, Andrew, Jonathan Dombrosky, Barney Venables, and Steve Wolverton. 2018. "Taphonomy and Negative Results: An Integrated Approach to Ceramic-Bound Protein Residue Analysis." *Journal of Archaeological Science* 94: 32–43.

Barker, Andrew, Barney Venables, Stanley M. Stevens Jr, Kent W. Seeley, Peggy Wang, and Steve Wolverton. 2012. "An Optimized Approach for Protein Residue Extraction and Identification from Ceramics after Cooking." *Journal of Archaeological Method and Theory* 19 (3): 407–39.

Barre, Annick, Els J. M. Van Damme, Mathias Simplicien, Hervé Benoist, and Pierre Rougé. 2020. "Are Dietary Lectins Relevant Allergens in Plant Food Allergy?" *Foods (Basel, Switzerland)* 9 (12). https://doi.org/10.3390/foods9121724.

Beniash, Elia. 2011. "Biominerals--Hierarchical Nanocomposites: The Example of Bone." *Wiley Interdisciplinary Reviews. Nanomedicine and Nanobiotechnology* 3 (1): 47–69.

Berthelot, Karine, Sophie Lecomte, Bénédicte Coulary-Salin, Ahmed Bentaleb, and Frédéric Peruch. 2016. "Hevea Brasiliensis Prohevein Possesses a Conserved C-Terminal Domain with Amyloid-like Properties in Vitro." *Biochimica et Biophysica Acta* 1864 (4): 388–99.

Bienert, Stefan, Andrew Waterhouse, Tjaart A. P. de Beer, Gerardo Tauriello, Gabriel Studer, Lorenza Bordoli, and Torsten Schwede. 2016. "The SWISS-MODEL Repository—new Features and Functionality." *Nucleic Acids Research* 45 (D1): D313–19.

Bislev, S. L., E. W. Deutsch, Z. Sun, T. Farrah, R. Aebersold, R. Moritz, E. Bendixen, and M. Codrea. 2012. "A Bovine Peptide Atlas of Milk and Mammary Gland Proteomes." *Proteomics* 18: 2895–99.

Bleasdale, Madeleine, Kristine K. Richter, Anneke Janzen, Samantha Brown, Ashley Scott, Jana Zech, Shevan Wilkin, et al. 2021. "Ancient Proteins Provide Evidence of Dairy Consumption in Eastern Africa." *Nature Communications* 12 (1): 632.

Bogaard, Amy, and Alan K. Outram. 2013. "Palaeodiet and Beyond: Stable Isotopes in Bioarchaeology." *World Archaeology* 45 (3): 333–37.

Boskey, Adele L., and Eduardo Villarreal-Ramirez. 2016. "Intrinsically Disordered Proteins and Biomineralization." *Matrix Biology: Journal of the International Society for Matrix Biology* 52-54 (May): 43–59.

Brandt, Luise Ørsted, and Ulla Mannering. 2021. "Taxonomic Identification of Danish Viking Age Shoes and Skin Objects by ZooMS (Zooarchaeology by Mass Spectrometry)." *Journal of Proteomics* 231: 104038.

Buckley, Michael, Matthew Collins, and Jane Thomas-Oates. 2008. "A Method of Isolating the Collagen (I) alpha2 Chain Carboxytelopeptide for Species Identification in Bone Fragments." *Analytical Biochemistry* 374 (2): 325–34.

Buckley, Michael, Matthew Collins, Jane Thomas-Oates, and Julie C. Wilson. 2009. "Species Identification by Analysis of Bone Collagen Using Matrix-Assisted Laser Desorption/ionisation Time-of-Flight Mass Spectrometry." *Rapid Communications in Mass Spectrometry: RCM* 23 (23): 3843–54.

Buckley, M., and C. Wadsworth. 2014. "Proteome Degradation in Ancient Bone: Diagenesis and Phylogenetic Potential." *Palaeogeography, Palaeoclimatology, Palaeoecology* 416: 69–79.

Buckley, Stephen, Donatella Usai, Tina Jakob, Anita Radini, and Karen Hardy. 2014. "Dental Calculus Reveals Unique Insights into Food Items, Cooking and Plant Processing in Prehistoric Central Sudan." *PloS One* 9 (7): e100808.

Bu, Guanhao, Yongkang Luo, Fusheng Chen, Kunlun Liu, and Tingwei Zhu. 2013. "Milk Processing as a Tool to Reduce Cow's Milk Allergenicity: A Mini-Review." *Dairy Science & Technology* 93 (3): 211–23.

Burger, J., M. Kirchner, B. Bramanti, W. Haak, and M. G. Thomas. 2007. "Absence of the Lactase-Persistence-Associated Allele in Early Neolithic Europeans." *Proceedings of the National Academy of Sciences of the United States of America* 104 (10): 3736–41.

Cabanes, Dan, and Ruth Shahack-Gross. 2015. "Understanding Fossil Phytolith Preservation: The Role of Partial Dissolution in Paleoecology and Archaeology." *PloS One* 10 (5): e0125532.

Cao, Yiping, and Raffaele Mezzenga. 2019. “Food Protein Amyloid Fibrils: Origin, Structure, Formation, Characterization, Applications and Health Implications.” *Advances in Colloid and Interface Science* 269: 334–56.

Cappellini, Enrico, Anthea Gentry, Eleftheria Palkopoulou, Yasuko Ishida, David Cram, Anna-Marie Roos, Mick Watson, et al. 2014. “Resolution of the Type Material of the Asian Elephant, Elephas Maximus Linnaeus, 1758 (Proboscidea, Elephantidae).” *Zoological Journal of the Linnean Society* 170 (1): 222–32.

Cappellini, Enrico, M. Thomas P. Gilbert, Filippo Geuna, Girolamo Fiorentino, Allan Hall, Jane Thomas-Oates, Peter D. Ashton, et al. 2010. “A Multidisciplinary Study of Archaeological Grape Seeds.” *Die Naturwissenschaften* 97 (2): 205–17.

Cappellini, Enrico, Frido Welker, Luca Pandolfi, Jazmín Ramos-Madrigal, Diana Samodova, Patrick L. Rüther, Anna K. Fotakis, et al. 2019. “Early Pleistocene Enamel Proteome from Dmanisi Resolves Stephanorhinus Phylogeny.” *Nature* 574 (7776): 103–7.

Carneiro, Karina M. M., Halei Zhai, Li Zhu, Jeremy A. Horst, Melody Sitlin, Mychi Nguyen, Martin Wagner, et al. 2016. “Amyloid-like Ribbons of Amelogenins in Enamel Mineralization.” *Scientific Reports* 6: 23105.

Carrer, Francesco, André Carlo Colonese, Alexandre Lucquin, Eduardo Petersen Guedes, Anu Thompson, Kevin Walsh, Thomas Reitmaier, and Oliver E. Craig. 2016. “Chemical Analysis of Pottery Demonstrates Prehistoric Origin for High-Altitude Alpine Dairying.” *PloS One* 11 (4): e0151442.

Castellanos, Milagros, Almudena Torres-Pardo, Rosa Rodríguez-Pérez, and María Gasset. 2018. “Amyloid Assembly Endows Gad M 1 with Biomineralization Properties.” *Biomolecules* 8 (1). https://doi.org/10.3390/biom8010013.

Chaneton, L., J. M. Pérez Sáez, and L. E. Bussmann. 2011. “Antimicrobial Activity of Bovine β-Lactoglobulin against Mastitis-Causing Bacteria.” *Journal of Dairy Science* 94 (1): 138–45.

Charlton, Sophy, Abigail Ramsøe, Matthew Collins, Oliver E. Craig, Roman Fischer, Michelle Alexander, and Camilla F. Speller. 2019. “New Insights into Neolithic Milk Consumption through Proteomic Analysis of Dental Calculus.” *Archaeological and Anthropological Sciences* 11 (11): 6183–96.

Chen, Fahu, Frido Welker, Chuan-Chou Shen, Shara E. Bailey, Inga Bergmann, Simon Davis, Huan Xia, et al. 2019. “A Late Middle Pleistocene Denisovan Mandible from the Tibetan Plateau.” *Nature* 569 (7756): 409–12.

Colgrave, Michelle Lisa. 2017. *Proteomics in Food Science: From Farm to Fork*. Academic Press.

Collins, Matthew J., Michael S. Riley, Angela M. Child, and Gordon Turner-Walker. 1995. “A Basic Mathematical Simulation of the Chemical Degradation of Ancient Collagen.” *Journal of Archaeological Science* 22 (2): 175–83.

Collins, M. J., C. M. Nielsen- Marsh, J. Hiller, C. I. Smith, J. P. Roberts, R. V. Prigodich, J. Csapo, A. R. Millard, and G. Turner-Walker. 2002. “The Survival of Organic Matter in Bone: A Review.” *Archaeometry* 44 (3): 383–94.

Copley, M. S., R. Berstan, S. N. Dudd, G. Docherty, A. J. Mukherjee, V. Straker, S. Payne, and R. P. Evershed. 2003. “Direct Chemical Evidence for Widespread Dairying in Prehistoric Britain.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (4): 1524–29.

Corr, Lorna T., Judith C. Sealy, Mark C. Horton, and Richard P. Evershed. 2005. “A Novel Marine Dietary Indicator Utilising Compound-Specific Bone Collagen Amino Acid δ13C Values of Ancient Humans.” *Journal of Archaeological Science* 32 (3): 321–30.

Costa, Joana, Caterina Villa, Kitty Verhoeckx, Tanja Cirkovic-Velickovic, Denise Schrama, Paola Roncada, Pedro M. Rodrigues, et al. 2021. “Are Physicochemical Properties Shaping the Allergenic Potency of Animal Allergens?” *Clinical Reviews in Allergy & Immunology*. https://doi.org/10.1007/s12016-020-08826-1.

Craig, Oliver E., Richard B. Allen, Anu Thompson, Rhiannon E. Stevens, Valerie J. Steele, and Carl Heron. 2012. “Distinguishing Wild Ruminant Lipids by Gas Chromatography/combustion/isotope Ratio Mass Spectrometry.” *Rapid Communications in Mass Spectrometry: RCM* 26 (19): 2359–64.

Craig, Oliver E., Val J. Steele, Anders Fischer, Sönke Hartz, Søren H. Andersen, Paul Donohoe, Aikaterini Glykou, et al. 2011. “Ancient Lipids Reveal Continuity in Culinary Practices across the Transition to Agriculture in Northern Europe.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (44): 17910–15.

Craig, Robertson, John P. Cortens, and Ronald C. Beavis. 2005. “The Use of Proteotypic Peptide Libraries for Protein Identification.” *Rapid Communications in Mass Spectrometry* 19 (13): 1844–50.

Crowther, Jennifer M., Geoffrey B. Jameson, Alison J. Hodgkinson, and Renwick C. J. Dobson. 2016. “Structure, Oligomerisation and Interactions of β-Lactoglobulin.” In *Milk Proteins - From Structure to Biological Properties and Health Aspects*, edited by I. Gigli. IntechOpen.

Danielsen, Heidi N., Susan H. Hansen, Florian-Alexander Herbst, Henrik Kjeldal, Allan Stensballe, Per H. Nielsen, and Morten S. Dueholm. 2017. “Direct Identification of Functional Amyloid Proteins by Label-Free Quantitative Mass Spectrometry.” *Biomolecules* 7 (3). https://doi.org/10.3390/biom7030058.

Day, Jo. 2013. “Botany Meets Archaeology: People and Plants in the Past.” *Journal of Experimental Botany* 64 (18): 5805–16.

De Baets, Greet, Joost Schymkowitz, and Frederic Rousseau. 2014. “Predicting Aggregation-Prone Sequences in Proteins.” *Essays in Biochemistry* 56: 41–52.

Demarchi, Beatrice. 2020. *Amino Acids and Proteins in Fossil Biominerals: An Introduction for Archaeologists and Palaeontologists*. John Wiley & Sons.

Demarchi, Beatrice, Shaun Hall, Teresa Roncal-Herrero, Colin L. Freeman, Jos Woolley, Molly K. Crisp, Julie Wilson, et al. 2016. “Protein Sequences Bound to Mineral Surfaces Persist into Deep Time.” *eLife* 5. https://doi.org/10.7554/eLife.17092.

Demarchi, Beatrice, Samantha Presslee, Jorune Sakalauskaite, Roman Fischer, and Julia Best. 2020. “The Role of Birds at Çatalhöyük Revealed by the Analysis of Eggshell.” *Quaternary International: The Journal of the International Union for Quaternary Research* 543: 50–60.

Demarchi, B., S. Presslee, and I. Gutiérrez-Zugasti. 2019. “Birds of Prey and Humans in Prehistoric Europe: A View from El Mirón Cave, Cantabria (Spain).” *Journal of Archaeological Science: Reports* 24 (330): 224–52.

Desmond, Abigail, Nick Barton, Abdeljalil Bouzouggar, Katerina Douka, Philippe Fernandez, Louise Humphrey, Jacob Morales, Elaine Turner, and Michael Buckley. 2018. “ZooMS Identification of Bone Tools from the North African Later Stone Age.” *Journal of Archaeological Science* 98: 149–57.

Dias, Renata de Oliveira, Leandro Dos Santos Machado, Ludovico Migliolo, and Octavio Luiz Franco. 2015. “Insights into Animal and Plant Lectins with Antimicrobial Activities.” *Molecules* 20 (1): 519–41.

Dobberstein, R. C., M. J. Collins, O. E. Craig, G. Taylor, K. E. H. Penkman, and S. Ritz-Timme. 2009. “Archaeological Collagen: Why Worry about Collagen Diagenesis?” *Archaeological and Anthropological Sciences* 1 (1): 31–42.

Domínguez-Rodrigo, Manuel. 2012. “Critical Review of the MNI (minimum Number of Individuals) as a Zooarchaeological Unit of Quantification.” *Archaeological and Anthropological Sciences* 4 (1): 47–59.

Dongoske, K. E., D. L. Martin, and T. J. Ferguson. 2000. “Critique of the Claim of Cannibalism at Cowboy Wash.” *American Antiquity* 65 (1): 179–90.

Downs, Elinor F., and Jerold M. Lowenstein. 1995. “Identification of Archaeological Blood Proteins: A Cautionarynote.” *Journal of Archaeological Science* 22 (1): 11–16.

Drieu, Léa, Cédric Lepère, and Martine Regert. 2020. “The Missing Step of Pottery Chaîne Opératoire: Considering Post-Firing Treatments on Ceramic Vessels Using Macro- and Microscopic Observation and Molecular Analysis.” *Journal of Archaeological Method and Theory*. https://doi.org/10.1007/s10816-019-09428-8.

Dubé, André, Cindy Leggiadro, and Kathryn Vanya Ewart. 2016. “Rapid Amyloid Fibril Formation by a Winter Flounder Antifreeze Protein Requires Specific Interaction with Ice.” *FEBS Letters* 590 (9): 1335–44.

Dunne, J., A. Chapman, P. Blinkhorn, and R. P. Evershed. 2020. “Fit for Purpose? Organic Residue Analysis and Vessel Specialisation: The Perfectly Utilitarian Medieval Pottery Assemblage from West Cotton, Raunds.” *Journal of Archaeological Science* 120: 105178.

Dunne, J., Richard P. Evershed, Mélanie Salque, Lucy Cramp, Silvia Bruni, Kathleen Ryan, Stefano Biagetti, and Savino di Lernia. 2012. “First Dairying in Green Saharan Africa in the Fifth Millennium BC.” *Nature* 486 (7403): 390–94.

Dziuba, Bartłomiej, and Marta Dziuba. 2014. “New Milk Protein-Derived Peptides with Potential Antimicrobial Activity: An Approach Based on Bioinformatic Studies.” *International Journal of Molecular Sciences* 15 (8): 14531–45.

Eerkens, J. W. 2005. “GC–MS Analysis and Fatty Acid Ratios of Archaeological Potsherds from the Western Great Basin of North America.” *Archaeometry* 47: 83–102.

Engeroff, Paul, and Monique Vogel. 2021. “The Role of CD23 in the Regulation of Allergic Responses.” *Allergy* 76 (7): 1981–89.

Eriksson, G. 2013. “Stable Isotope Analysis of Humans.” In *The Oxford Handbook of the Archaeology of Death and Burial*, edited by L. Nilsson Stutz and S. Tarlow, 123–46. Oxford University Press.

Evans, John G., Terence Patrick O’Connor, and Terry O’Connor. 1999. *Environmental Archaeology: Principles and Methods*. Sutton Pub Limited.

Evans, John Spencer. 2020. “Glycosylation: A ‘Last Word’ in the Protein-Mediated Biomineralization Process.” *Crystals* 10 (9): 818.

Evershed, Richard P., Sebastian Payne, Andrew G. Sherratt, Mark S. Copley, Jennifer Coolidge, Duska Urem-Kotsu, Kostas Kotsakis, et al. 2008. “Earliest Date for Milk Use in the Near East and Southeastern Europe Linked to Cattle Herding.” *Nature* 455 (7212): 528–31.

Evershed, R. P. 2008. “Organic Residue Analysis in Archaeology: The Archaeological Biomarker Revolution.” *Archaeometry* 50 (6): 895–924.

Fernandes, Ricardo, Yvette Eley, Marek Brabec, Alexandre Lucquin, Andrew Millard, and Oliver E. Craig. 2018. “Reconstruction of Prehistoric Pottery Use from Fatty Acid Carbon Isotope Signatures Using Bayesian Inference.” *Organic Geochemistry* 117: 31–42.

Fernandes, Ricardo, and Klervia Jaouen. 2017. “Isotopes in Archaeology.” *Archaeological and Anthropological Sciences* 9 (7): 1305–6.

Fernandes, Ricardo, Andrew R. Millard, Marek Brabec, Marie-Josée Nadeau, and Pieter Grootes. 2014. “Food Reconstruction Using Isotopic Transferred Signals (FRUITS): A Bayesian Model for Diet Reconstruction.” *PloS One* 9 (2): e87436.

Fiddyment, Sarah, Bruce Holsinger, Chiara Ruzzier, Alexander Devine, Annelise Binois, Umberto Albarella, Roman Fischer, et al. 2015. “Animal Origin of 13th-Century Uterine Vellum Revealed Using Noninvasive Peptide Fingerprinting.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (49): 15066–71.

Fiedel, Stuart. 1996. “Blood from Stones? Some Methodological and Interpretive Problems in Blood Residue Analysis.” *Journal of Archaeological Science* 23 (1): 139–47.

Fogel, Marilyn L., Noreen Tuross, Beverly J. Johnson, and Gifford H. Miller. 1997. “Biogeochemical Record of Ancient Humans.” *Organic Geochemistry* 27 (5): 275–87.

Fuss, Jochen, Gregor Uhlig, and Madelaine Böhme. 2018. “Earliest Evidence of Caries Lesion in Hominids Reveal Sugar-Rich Diet for a Middle Miocene Dryopithecine from Europe.” *PloS One* 13 (8): e0203307.

Gamble, Clive. 1978. “Optimising Information from Studies of Faunal Remains.” In *Sampling in Contemporary British Archaeology*, edited by John Cherry, Clive Gamble, and Stephen Shennan, 411. 50. British Archaeological Reports.

Garvey, Megan, Sarah Meehan, Sally L. Gras, Horst J. Schirra, David J. Craik, Nicole L. Van der Weerden, Marilyn A. Anderson, Juliet A. Gerrard, and John A. Carver. 2013. “A Radish Seed Antifungal Peptide with a High Amyloid Fibril-Forming Propensity.” *Biochimica et Biophysica Acta* 1834 (8): 1615–23.

Gattinger, Pia, Urska Bidovec-Stojkovic, Mihaela Zidarn, Peter Korosec, Rudolf Valenta, and Irene Mittermann. 2021. “Glycosylation Enhances Allergenic Activity of Major Bee Venom Allergen Api M 1 by Adding IgE Epitopes.” *The Journal of Allergy and Clinical Immunology* 147 (4): 1502–4.e5.

Geber, Jonny, Monica Tromp, Ashley Scott, Abigail Bouwman, Paolo Nanni, Jonas Grossmann, Jessica Hendy, and Christina Warinner. 2019. "Relief Food Subsistence Revealed by Microparticle and Proteomic Analyses of Dental Calculus from Victims of the Great Irish Famine." *Proceedings of the National Academy of Sciences* 116 (39): 19380–85.

Gerbault, Pascale, Anke Liebert, Yuval Itan, Adam Powell, Mathias Currat, Joachim Burger, Dallas M. Swallow, and Mark G. Thomas. 2011. "Evolution of Lactase Persistence: An Example of Human Niche Construction." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 366 (1566): 863–77.

Goldschmidt, L., P. K. Teng, and R. Riek. 2010. "Identifying the Amylome, Proteins Capable of Forming Amyloid-like Fibrils." *Proceedings of the National Academy of Sciences.* 107 (8) 3487-3492

Gour, S., V. Kaushik, V. Kumar, and P. Bhat. 2016. "Antimicrobial Peptide (Cn-AMP2) from Liquid Endosperm of Cocos Nucifera Forms Amyloid-like Fibrillar Structure." *Journal of Peptide Science* 22: 201– 207. doi: 10.1002/psc.2860.

Graether, S. P., and B. D. Sykes. 2004. "Cold Survival in Freeze-intolerant Insects: The Structure and Function of β-helical Antifreeze Proteins." *European Journal of Biochemistry / FEBS* 217 (16): 3285–96.

Grillo, Katherine M., Julie Dunne, Fiona Marshall, Mary E. Prendergast, Emmanuelle Casanova, Agness O. Gidna, Anneke Janzen, et al. 2020. "Molecular and Isotopic Evidence for Milk, Meat, and Plants in Prehistoric Eastern African Herder Food Systems." *Proceedings of the National Academy of Sciences of the United States of America* 117 (18): 9793–99.

Grzybowska, Ewa A. 2018. "Calcium-Binding Proteins with Disordered Structure and Their Role in Secretion, Storage, and Cellular Signaling." *Biomolecules* 8 (2). https://doi.org/10.3390/biom8020042.

Guenther, Elizabeth L., Peng Ge, Hamilton Trinh, Michael R. Sawaya, Duilio Cascio, David R. Boyer, Tamir Gonen, Z. Hong Zhou, and David S. Eisenberg. 2018. "Atomic-Level Evidence for Packing and Positional Amyloid Polymorphism by Segment from TDP-43 RRM2." *Nature Structural & Molecular Biology* 25 (4): 311–19.

Haldane, Cheryl. 1993. "Direct Evidence for Organic Cargoes in the Late Bronze Age." *World Archaeology* 24 (3): 348–60.

Haslam, Michael, Gail Robertson, Alison Crowther, Sue Nugent, and Luke Kirkwood. 2009. *Archaeological Science Under a Microscope: Studies in Residue and Ancient DNA Analysis in Honour of Thomas H. Loy*. ANU E Press.

Havlis, Jan, Henrik Thomas, Marek Sebela, and Andrej Shevchenko. 2003. "Fast-Response Proteomics by Accelerated in-Gel Digestion of Proteins." *Analytical Chemistry* 75 (6): 1300–1306.

Hendy, Jessica. 2021. "Ancient Protein Analysis in Archaeology." *Science Advances* 7 (3). https://doi.org/10.1126/sciadv.abb9314.

Hendy, Jessica, Andre C. Colonese, Ingmar Franz, Ricardo Fernandes, Roman Fischer, David Orton, Alexandre Lucquin, et al. 2018. "Ancient Proteins from Ceramic Vessels at Çatalhöyük West Reveal the Hidden Cuisine of Early Farmers." *Nature Communications* 9 (1): 4064.

Hendy, Jessica, Christina Warinner, Abigail Bouwman, Matthew J. Collins, Sarah Fiddyment, Roman Fischer, Richard Hagan, et al. 2018. "Proteomic Evidence of Dietary Sources in Ancient Dental Calculus." *Proceedings. Biological Sciences / The Royal Society* 285 (1883). https://doi.org/10.1098/rspb.2018.0977.

Henry, Amanda G. 2020a. "Starch Granules as Markers of Diet and Behavior." In *Handbook for the Analysis of Micro-Particles in Archaeological Samples*, edited by Amanda G. Henry, 97–116. Cham: Springer International Publishing.

———. 2020b. *Handbook for the Analysis of Micro-Particles in Archaeological Samples*. Springer Nature.

Henry, Amanda G., and Dolores R. Piperno. 2008. "Using Plant Microfossils from Dental Calculus to Recover Human Diet: A Case Study from Tell Al-Raqā'i, Syria." *Journal of Archaeological Science* 35 (7): 1943–50.

Hillman, Gordon C. 1986. “Plant Foods in Ancient Diet: The Archaeological Role of Palaeofaeces in General and Lindow Man’s Gut Contents in Particular.” *Lindow Man: The Body in the Bog*, 90–114.

Hong, Chuan, Hongen Jiang, Enguo Lü, Yunfei Wu, Lihai Guo, Yongming Xie, Changsui Wang, and Yimin Yang. 2012. “Identification of Milk Component in Ancient Food Residue by Proteomics.” *PloS One* 7 (5): e37053.

Hou, Juncai, Zhijing Liu, Songsong Cao, Haimei Wang, Chenggang Jiang, Muhammad Altaf Hussain, and Shiyue Pang. 2018. “Broad-Spectrum Antimicrobial Activity and Low Cytotoxicity against Human Cells of a Peptide Derived from Bovine αS1-Casein.” *Molecules* 23 (5). https://doi.org/10.3390/molecules23051220.

Hunt, C. O., G. Rushworth, D. D. Gilbertson, and D. J. Mattingly. 2001. “Romano-Libyan Dryland Animal Husbandry and Landscape: Pollen and Palynofacies Analyses of Coprolites from a Farm in the Wadi El-Amud, Tripolitania.” *Journal of Archaeological Science* 28 (4): 351–63.

Hutschenreuther, Antje, Jörg Watzke, Simone Schmidt, Thomas Büdel, and Amanda G. Henry. 2017. “Archaeological Implications of the Digestion of Starches by Soil Bacteria: Interaction among Starches Leads to Differential Preservation.” *Journal of Archaeological Science: Reports* 15: 95–108.

Iannuzzi, Clara, Gaetano Irace, and Ivana Sirangelo. 2015. “The Effect of Glycosaminoglycans (GAGs) on Amyloid Aggregation and Toxicity.” *Molecules* 20 (2): 2510–28.

Iconomidou, V. A., G. Vriend, and S. J. Hamodrakas. 2000. “Amyloids Protect the Silkmoth Oocyte and Embryo.” *FEBS Letters* 479 (3): 141–45.

Irving-Pease, Evan K., Hannah Ryan, Alexandra Jamieson, Evangelos A. Dimopoulos, Greger Larson, and Laurent A. F. Frantz. 2019. “Paleogenomics of Animal Domestication.” In *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, edited by Charlotte Lindqvist and Om P. Rajora, 225–72. Cham: Springer International Publishing.

Jacob, Reeba S., Edna George, Pradeep K. Singh, Shimul Salot, Arunagiri Anoop, Narendra Nath Jha, Shamik Sen, and Samir K. Maji. 2016. “Cell Adhesion on Amyloid Fibrils Lacking Integrin Recognition Motif.” *The Journal of Biological Chemistry* 291 (10): 5278–98.

Jakubovics, Nicholas S., Steven D. Goodman, Lauren Mashburn-Warren, Graham P. Stafford, and Fabian Cieplik. 2021. “The Dental Plaque Biofilm Matrix.” *Periodontology 2000* 86 (1): 32–56.

Jansens, K. J. A., I. Rombouts, C. Grootaert, K. Brijs, J. Van Camp, P. Van der Meeren, F. Rousseau, Schymkowitz J., and J. A. Delcour. 2019. “Rational Design of Amyloid-like Fibrillary Structures for Tailoring Food Protein Techno-functionality and Their Potential Health Implications.” *Comprehensive Reviews in Food Science and Food Safety* 18: 84–105.

Jansens, Koen J. A., Marlies A. Lambrecht, Ine Rombouts, Margarita Monge Morera, Kristof Brijs, Frederic Rousseau, Joost Schymkowitz, and Jan A. Delcour. 2019. “Conditions Governing Food Protein Amyloid Fibril Formation-Part I: Egg and Cereal Proteins.” *Comprehensive Reviews in Food Science and Food Safety* 18 (4): 1256–76.

Janzen, Anneke, Kristine Korzow Richter, Ogeto Mwebi, Samantha Brown, Veronicah Onduso, Filia Gatwiri, Emmanuel Ndiema, 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.

Jarnuczak, Andrew F., Dave C. H. Lee, Craig Lawless, Stephen W. Holman, Claire E. Eyers, and Simon J. Hubbard. 2016. “Analysis of Intrinsic Peptide Detectability via Integrated Label-Free and SRM-Based Absolute Quantitative Proteomics.” *Journal of Proteome Research* 15 (9): 2945–59.

Jeong, Choongwon, Ke Wang, Shevan Wilkin, William Timothy Treal Taylor, Bryan K. Miller, Jan H. Bemmann, Raphaela Stahl, et al. 2020. “A Dynamic 6,000-Year Genetic History of Eurasia’s Eastern Steppe.” *Cell* 183 (4): 890–904.e29.

Jeong, Choongwon, Shevan Wilkin, Tsend Amgalantugs, Abigail S. Bouwman, William Timothy Treal Taylor, Richard W. Hagan, Sabri Bromage, et al. 2018. “Bronze Age Population Dynamics and the Rise of Dairy Pastoralism on the Eastern Eurasian Steppe.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (48): E11248–55.

Jersie-Christensen, Rosa R., Liam T. Lanigan, David Lyon, Meaghan Mackie, Daniel Belstrøm, Christian D. Kelstrup, Anna K. Fotakis, et al. 2018. “Quantitative Metaproteomics of Medieval Dental Calculus Reveals Individual Oral Health Status.” *Nature Communications* 9 (1): 4744.

Katzenberg, M. A., and A. Grauer. 2018. “Stable Isotope Analysis: A Tool for Studying Past Diet, Demography, and Life History.” In *Biological Anthropology of the Human Skeleton*, edited by M. A. Katzenberg and A. Grauer, 469–504. Wiley-Blackwell.

Ke, Pu Chun, Ruhong Zhou, Louise C. Serpell, Roland Riek, Tuomas P. J. Knowles, Hilal A. Lashuel, Ehud Gazit, et al. 2020. “Half a Century of Amyloids: Past, Present and Future.” *Chemical Society Reviews* 49 (15): 5473–5509.

Kirby, Daniel P., Michael Buckley, Ellen Promise, Sunia A. Trauger, and T. Rose Holdcraft. 2013. “Identification of Collagen-Based Materials in Cultural Heritage.” *The Analyst* 138 (17): 4849–58.

Knipper, Corina, Sabine Reinhold, Julia Gresky, Nataliya Berezina, Claudia Gerling, Sandra L. Pichler, Alexandra P. Buzhilova, et al. 2020. “Diet and Subsistence in Bronze Age Pastoral Communities from the Southern Russian Steppes and the North Caucasus.” *PloS One* 15 (10): e0239861.

Kollipara, Laxmikanth, and René P. Zahedi. 2013. “Protein Carbamylation: In Vivo Modification or in Vitro Artefact?” *Proteomics* 13 (6): 941–44.

Korzow Richter, Kristine, Krista McGrath, Edouard Masson-MacLean, Simon Hickinbotham, Andrew Tedder, Kate Britton, Zoe Bottomley, et al. 2020. “What’s the Catch? Archaeological Application of Rapid Collagen-Based Species Identification for Pacific Salmon.” *Journal of Archaeological Science* 116: 105116.

Krueger, H. W., and C. H. Sullivan. 1984. “Models for Carbon Isotope Fractionation between Diet and Bone.” In *Stable Isotopes in Nutrition*, edited by Turnlund et al., 205–20. ACS Symposium Series. ACS Publications.

Lakshminarayanan, R., and R. M. Kini. 2002. “Investigation of the Role of Ansocalcin in the Biomineralization in Goose Eggshell Matrix.” *Proceedings of the National Academy of Sciences* 99 (8) 5155-5159; DOI:10.1073/pnas.072658899

Lambacher, Nicole, Karina Gerdau-Radonic, Emma Bonthorne, and Francisco José Valle de Tarazaga Montero. 2016. “Evaluating Three Methods to Estimate the Number of Individuals from a Commingled Context.” *Journal of Archaeological Science: Reports* 10: 674–83.

Lambrecht, Marlies A., Koen J. A. Jansens, Ine Rombouts, Kristof Brijs, Frederic Rousseau, Joost Schymkowitz, and Jan A. Delcour. 2019. “Conditions Governing Food Protein Amyloid Fibril Formation. Part II: Milk and Legume Proteins.” *Comprehensive Reviews in Food Science and Food Safety* 18 (4): 1277–91.

Lambrecht, Marlies A., Margarita Monge-Morera, Thibault Godefroidt, Nele Vluymans, Lomme J. Deleu, Peter Goos, Joost Schymkowitz, Frederic Rousseau, and Jan A. Delcour. 2021. “Hydrothermal Treatments Cause Wheat Gluten-Derived Peptides to Form Amyloid-like Fibrils.” *Journal of Agricultural and Food Chemistry* 69 (6): 1963–74.

Lee Lyman, R., and Cma Lyman. 1994. *Vertebrate Taphonomy*. Cambridge University Press.

Lee-Thorp, Julia, and Matt Sponheimer. 2003. “Three Case Studies Used to Reassess the Reliability of Fossil Bone and Enamel Isotope Signals for Paleodietary Studies.” *Journal of Anthropological Archaeology* 22 (3): 208–16.

León, Ileana R., Veit Schwämmle, Ole N. Jensen, and Richard R. Sprenger. 2013. “Quantitative Assessment of in-Solution Digestion Efficiency Identifies Optimal Protocols for Unbiased Protein Analysis.” *Molecular & Cellular Proteomics: MCP* 12 (10): 2992–3005.

Lermyte, Frederik. 2020. “Roles, Characteristics, and Analysis of Intrinsically Disordered Proteins: A Minireview.” *Life* 10 (12). https://doi.org/10.3390/life10120320.

Lieverse, Angela R. 1999. “Diet and the Aetiology of Dental Calculus.” *International Journal of Osteoarchaeology* 9 (4): 219–32.

Li, Li, Lei Zhu, and Yaoting Xie. 2021. “Proteomics Analysis of the Soil Textile Imprints from Tomb M6043 of the Dahekou Cemetery Site in Yicheng County, Shanxi Province, China.” *Archaeological and Anthropological Sciences* 13 (1). https://doi.org/10.1007/s12520-020-01258-0.

Liu, Xiaojun, Shaojian Dong, Can Jin, Zhiyi Bai, Guiling Wang, and Jiale Li. 2015. "Silkmapin of Hyriopsis Cumingii, a Novel Silk-like Shell Matrix Protein Involved in Nacre Formation." *Gene* 555 (2): 217–22.

Liu, Xiaojun, Can Jin, Leiming Wu, Shaojian Dong, Shimei Zeng, and Jiale Li. 2017. "Hic74, a Novel Alanine and Glycine Rich Matrix Protein Related to Nacreous Layer Formation in the Mollusc Hyriopsis Cumingii." *Aquaculture and Fisheries* 2 (3): 119–23.

Liu, X., J. Pu, S. Zeng, C. Jin, S. Dong, and J. Li. 2017. "Hyriopsis Cumingii Hic52—a Novel Nacreous Layer Matrix Protein with a Collagen-like Structure." *International Journal of Biological Macromolecules* 102: 667–73.

Loveday, Simon M., Skelte G. Anema, and Harjinder Singh. 2017. "β-Lactoglobulin Nanofibrils: The Long and the Short of It." *International Dairy Journal / Published in Association with the International Dairy Federation* 67: 35–45.

Lovell, N. C., B. S. Chisholm, D. E. Nelson, and H. P. Schwarcz. 1986. "Prehistoric Salmon Consumption in Interior British Columbia." *Canadian Journal of Archaeology / Journal Canadien d'Archéologie* 10: 99–106.

Mahajan, Sangeeta, and Vijay Sathe. 2020. "Scope and Limitations of Stable Isotope Analysis in Indian Archaeology." *Ancient Asia* 11 (0): 6.

Maixner, Frank, Dmitrij Turaev, Amaury Cazenave-Gassiot, Marek Janko, Ben Krause-Kyora, Michael R. Hoopmann, Ulrike Kusebauch, et al. 2018. "The Iceman's Last Meal Consisted of Fat, Wild Meat, and Cereals." *Current Biology*. https://doi.org/10.1016/j.cub.2018.05.067.

Mann, Allison E., James A. Fellows Yates, Zandra Fagernäs, Rita M. Austin, Elizabeth A. Nelson, and Courtney A. Hofman. 2020. "Do I Have Something in My Teeth? The Trouble with Genetic Analyses of Diet from Archaeological Dental Calculus." *Quaternary International.* https://doi.org/10.1016/j.quaint.2020.11.019.

Mann, Karlheinz, and Frank Siedler. 2004. "Ostrich (Struthio Camelus) Eggshell Matrix Contains Two Different C-Type Lectin-like Proteins. Isolation, Amino Acid Sequence, and Posttranslational Modifications." *Biochimica et Biophysica Acta* 1696 (1): 41–50.

Marciniak, Arkadiusz. 2011. "The Secondary Products Revolution: Empirical Evidence and Its Current Zooarchaeological Critique." *Journal of World Prehistory* 24 (2-3): 117–30.

Marshall, John D., J. Renee Brooks, and Kate Lajtha. 2007. "Sources of Variation in the Stable Isotopic Composition of Plants." In *Stable Isotopes in Ecology and Environmental Science*, edited by R. Michener and K. Lajtha, 22–60. John Wiley & Sons.

Martínez, Javier, Rosa Sánchez, Milagros Castellanos, Ana M. Fernández-Escamilla, Sonia Vázquez-Cortés, Montserrat Fernández-Rivas, and María Gasset. 2015. "Fish β-Parvalbumin Acquires Allergenic Properties by Amyloid Assembly." *Swiss Medical Weekly* 145: w14128.

Martisius, Naomi L., Frido Welker, Tamara Dogandžić, Mark N. Grote, William Rendu, Virginie Sinet-Mathiot, Arndt Wilcke, Shannon J. P. McPherron, Marie Soressi, and Teresa E. Steele. 2020. "Non-Destructive ZooMS Identification Reveals Strategic Bone Tool Raw Material Selection by Neandertals." *Scientific Reports* 10 (1): 7746.

McGovern, Patrick E., Thomas L. Sever, J. Wilson Myers, Eleanor Emlen Myers, Bruce Bevan, Naomi F. Miller, S. Bottema, et al. 1995. "Science in Archaeology: A Review." *American Journal of Archaeology* 99 (1): 79–142.

McGrath, Krista, Keri Rowsell, Christian Gates St-Pierre, Andrew Tedder, George Foody, Carolynne Roberts, Camilla Speller, and Matthew Collins. 2019. "Identifying Archaeological Bone via Non-Destructive ZooMS and the Materiality of Symbolic Expression: Examples from Iroquoian Bone Points." *Scientific Reports* 9 (1): 11027.

Mercader, Julio, Matthew Abtosway, Robert Bird, Mariam Bundala, Siobhan Clarke, Julien Favreau, Jamie L. Inwood, et al. 2018. "Morphometrics of Starch Granules from Sub-Saharan Plants and the Taxonomic Identification of Ancient Starch." *Frontiers of Earth Science in China* 6: 146.

Merwe, N. J. van der, and J. C. Vogel. 1978. "13C Content of Human Collagen as a Measure of Prehistoric Diet in Woodland North America." *Nature* 276 (5690): 815–16.

Michalski, Annette, Juergen Cox, and Matthias Mann. 2011. "More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority Is Inaccessible to Data-Dependent LC-MS/MS." *Journal of Proteome Research* 10 (4): 1785–93.

Monge-Morera, Margarita, Marlies A. Lambrecht, Lomme J. Deleu, Rodrigo Gallardo, Nikolaos N. Louros, Matthias De Vleeschouwer, Frederic Rousseau, Joost Schymkowitz, and Jan A. Delcour. 2020. "Processing Induced Changes in Food Proteins: Amyloid Formation during Boiling of Hen Egg White." *Biomacromolecules* 21 (6): 2218–28.

Monge-Morera, Margarita, Marlies A. Lambrecht, Lomme J. Deleu, Thibault Godefroidt, Peter Goos, Frederic Rousseau, Joost Schymkowitz, and Jan A. Delcour. 2021. "Drying Mode and Hydrothermal Treatment Conditions Govern the Formation of Amyloid-like Protein Fibrils in Solutions of Dried Hen Egg White." *Food Hydrocolloids* 112: 106276.

Monge-Morera, Margarita, Marlies A. Lambrecht, Lomme J. Deleu, Nikolaos N. Louros, Frederic Rousseau, Joost Schymkowitz, and Jan A. Delcour. 2021. "Heating Wheat Gluten Promotes the Formation of Amyloid-like Fibrils." *ACS Omega* 6 (3): 1823–33.

Naganuma, Takako, Wataru Hoshino, Yukihiro Shikanai, Rie Sato, Kaiyue Liu, Saho Sato, Koji Muramoto, Makoto Osada, Kyosuke Yoshimi, and Tomohisa Ogawa. 2014. "Novel Matrix Proteins of Pteria Penguin Pearl Oyster Shell Nacre Homologous to the Jacalin-Related β-Prism Fold Lectins." *PloS One* 9 (11): e112326.

Nelson, Rebecca, Michael R. Sawaya, Melinda Balbirnie, Anders Ø. Madsen, Christian Riekel, Robert Grothe, and David Eisenberg. 2005. "Structure of the Cross-β Spine of Amyloid-like Fibrils." *Nature* 435 (7043): 773–78.

Nielsen-Marsh, Christina M., Christin Stegemann, Ralf Hoffmann, Tanya Smith, Robin Feeney, Michel Toussaint, Katerina Harvati, Eleni Panagopoulou, Jean-Jacques Hublin, and Michael P. Richards. 2009. "Extraction and Sequencing of Human and Neanderthal Mature Enamel Proteins Using MALDI-TOF/TOF MS." *Journal of Archaeological Science* 36 (8): 1758–63.

Nistelberger, H. M., O. Smith, N. Wales, B. Star, and S. Boessenkool. 2016. "The Efficacy of High-Throughput Sequencing and Target Enrichment on Charred Archaeobotanical Remains." *Scientific Reports* 6: 37347.

Oli, M. W., H. N. Otoo, P. J. Crowley, and K. P. Heim. 2012. "Functional Amyloid Formation by Streptococcus Mutans." *Microbiology* 158: 2903–2916.

Omasits, Ulrich, Christian H. Ahrens, Sebastian Müller, and Bernd Wollscheid. 2013. "Protter: Interactive Protein Feature Visualization and Integration with Experimental Proteomic Data." *Bioinformatics* 30 (6): 884–86.

Ottoni, Claudio, Dušan Borić, Olivia Cheronet, Vitale Sparacello, Irene Dori, Alfredo Coppa, Dragana Antonović, et al. 2021. "Tracking the Transition to Agriculture in Southern Europe through Ancient DNA Analysis of Dental Calculus." *Proceedings of the National Academy of Sciences* 118 (32). https://doi.org/10.1073/pnas.2102116118.

Ozcan, Sureyya, Bum Jin Kim, Grace Ro, Jae-Han Kim, Thomas L. Bereuter, Christian Reiter, Lauren Dimapasoc, et al. 2014. "Glycosylated Proteins Preserved over Millennia: N-Glycan Analysis of Tyrolean Iceman, Scythian Princess and Warrior." *Scientific Reports* 4: 4963.https://doi.org/10.1038/srep04963

Pal, Prabir, Mrityunjoy Mahato, Tapanendu Kamilya, Bidisha Tah, Ratan Sarkar, and G. B. Talapatra. 2011. "Fibrillation of Egg White Ovalbumin: A Pathway via Biomineralization." *The Journal of Physical Chemistry. B* 115 (14): 4259–65.

Parker, G. J., J. M. Yip, J. W. Eerkens, and M. Salemi. 2019. "Sex Estimation Using Sexually Dimorphic Amelogenin Protein Fragments in Human Enamel." *Journal of Archaeological Science (2019).* 101:169-180.https://doi.org/10.1016/j.jas.2018.08.011.

Parker, Glendon J., Julia M. Yip, Jelmer W. Eerkens, Michelle Salemi, Blythe Durbin-Johnson, Caleb Kiesow, Randall Haas, et al. 2019. "Sex Estimation Using Sexually Dimorphic Amelogenin Protein Fragments in Human Enamel." *Journal of Archaeological Science* 101: 169–80.

Pekar, Judith, Davide Ret, and Eva Untersmayr. 2018. "Stability of Allergens." *Molecular Immunology* 100: 14–20.

Pentony, Melissa M., and David T. Jones. 2010. "Modularity of Intrinsic Disorder in the Human Proteome." *Proteins* 78 (1): 212–21.

Peres, Tanya M. 2010. "Methodological Issues in Zooarchaeology." In *Integrating Zooarchaeology and Paleoethnobotany: A Consideration of Issues, Methods, and Cases*, edited by Amber M. Van Der warker and Tanya M. Peres, 15–36. New York, NY: Springer New York.

Pérez-Tavarez, Raquel, Helena M. Moreno, Javier Borderias, David Loli-Ausejo, María Pedrosa, José Luis Hurtado, Rosa Rodriguez-Pérez, and María Gasset. 2021. “Fish Muscle Processing into Seafood Products Reduces β-Parvalbumin Allergenicity.” *Food Chemistry* 364: 130308.

Petroviciu, Irina, Florin Albu, and Andrei Medvedovici. 2010. “LC/MS and LC/MS/MS Based Protocol for Identification of Dyes in Historic Textiles.” *Microchemical Journal, Devoted to the Application of Microtechniques in All Branches of Science* 95 (2): 247–54.

Petzold, Axel, Ching-Hua Lu, Mike Groves, Johan Gobom, Henrik Zetterberg, Gerry Shaw, and Sonia O’Connor. 2020. “Protein Aggregate Formation Permits Millennium-Old Brain Preservation.” *Journal of the Royal Society, Interface / the Royal Society* 17 (162): 20190775.

Piperno, Dolores R. 2006. *Phytoliths: A Comprehensive Guide for Archaeologists and Paleoecologists*. Rowman Altamira.

Pont, Caroline, Stefanie Wagner, Antoine Kremer, Ludovic Orlando, Christophe Plomion, and Jerome Salse. 2019. “Paleogenomics: Reconstruction of Plant Evolutionary Trajectories from Modern and Ancient DNA.” *Genome Biology* 20 (1): 29.

Presslee, Samantha, Julie Wilson, Jos Woolley, Julia Best, Douglas Russell, Anita Radini, Roman Fischer, et al. 2017. “The Identification of Archaeological Eggshell Using Peptide Markers.” *STAR: Science & Technology of Archaeological Research* 3 (1): 89–99.

Radini, Anita, Efthymia Nikita, Stephen Buckley, Les Copeland, and Karen Hardy. 2017. “Beyond Food: The Multiple Pathways for Inclusion of Materials into Ancient Dental Calculus.” *American Journal of Physical Anthropology* 162 Suppl 63: 71–83.

Radini, A., M. Tromp, A. Beach, E. Tong, C. Speller, M. McCormick, J. V. Dudgeon, et al. 2019. “Medieval Women’s Early Involvement in Manuscript Production Suggested by Lapis Lazuli Identification in Dental Calculus.” *Science Advances* 5 (1): eaau7126.

Roffet-Salque, Mélanie, Julie Dunne, David T. Altoft, Emmanuelle Casanova, Lucy J. E. Cramp, Jessica Smyth, Helen L. Whelton, and Richard P. Evershed. 2017. “From the inside out: Upscaling Organic Residue Analyses of Archaeological Ceramics.” *Journal of Archaeological Science: Reports* 16: 627–40.

Ruiz-Arellano, R. R., F. J. Medrano, and A. Moreno. 2015. “Structure of Struthiocalcin-1, an Intramineral Protein from Struthio Camelus Eggshell, in Two Crystal Forms.” *Acta Crystallographica* 71: 809–18.

Russell, Nerissa. 2011. *Social Zooarchaeology: Humans and Animals in Prehistory*. Cambridge University Press.

Rybczynski, Natalia, John C. Gosse, C. Richard Harington, Roy A. Wogelius, Alan J. Hidy, and Mike Buckley. 2013. “Mid-Pliocene Warm-Period Deposits in the High Arctic Yield Insight into Camel Evolution.” *Nature Communications* 4: 1550.

Sakalauskaite, Jorune, Søren H. Andersen, Paolo Biagi, Maria A. Borrello, Théophile Cocquerez, André Carlo Colonese, Federica Dal Bello, et al. 2019. “‘Palaeoshellomics’ Reveals the Use of Freshwater Mother-of-Pearl in Prehistory.” *eLife* 8. https://doi.org/10.7554/eLife.45644.

Salque, Mélanie, Peter I. Bogucki, Joanna Pyzel, Iwona Sobkowiak-Tabaka, Ryszard Grygiel, Marzena Szmyt, and Richard P. Evershed. 2013. “Earliest Evidence for Cheese Making in the Sixth Millennium BC in Northern Europe.” *Nature* 493 (7433): 522–25.

Sánchez, Rosa, Javier Martínez, Ana Castro, María Pedrosa, Santiago Quirce, Rosa Rodríguez-Pérez, and María Gasset. 2016. “The Amyloid Fold of Gad M 1 Epitopes Governs IgE Binding.” *Scientific Reports* 6: 32801.

Santos, Jaime, and Salvador Ventura. 2021. “Functional Amyloids Germinate in Plants.” *Trends in Plant Science*. 27 (1):7-10. https://doi.org/10.1016/j.tplants.2020.10.001.

Sawafuji, Rikai, Aiko Saso, Wataru Suda, Masahira Hattori, and Shintaroh Ueda. 2020. “Ancient DNA Analysis of Food Remains in Human Dental Calculus from the Edo Period, Japan.” *PloS One* 15 (3): e0226654.

Sawaya, M. R., S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, and M. I. Apostol. 2007. “Atomic Structures of Amyloid Cross-Beta Spines Reveal Varied Steric Zippers.” *Nature* 447 (7143): 453–57.

Scott, Ashley, Robert C. Power, Victoria Altmann-Wendling, Michal Artzy, Mario A. S. Martin, Stefanie Eisenmann, Richard Hagan, et al. 2021. “Exotic Foods Reveal Contact between South

Asia and the Near East during the Second Millennium BCE." *Proceedings of the National Academy of Sciences.* 118 (2). https://doi.org/10.1073/pnas.2014956117.

Sealy, J. 2001. "Body Tissue Chemistry and Palaeodiet." In *Handbook of Archaeological Sciences,* edited by D. R. Brothwell and A. M. Pollard, 269–79. John Wiley & Sons.

Ségurel, Laure, and Céline Bon. 2017. "On the Evolution of Lactase Persistence in Humans." *Annual Review of Genomics and Human Genetics* 18 (August): 297–319.

Shoulders, Matthew D., and Ronald T. Raines. 2009. "Collagen Structure and Stability." *Annual Review of Biochemistry* 78 (1): 929–58.

Smith, Jeffrey F., Tuomas P. J. Knowles, Christopher M. Dobson, Cait E. Macphee, and Mark E. Welland. 2006. "Characterization of the Nanoscale Properties of Individual Amyloid Fibrils." *Proceedings of the National Academy of Sciences of the United States of America* 103 (43): 15806–11.

Solomon, A., T. Richey, C. L. Murphy, D. T. Weiss, J. S. Wall, G. T. Westermark, and P. Westermark. 2007. "Amyloidogenic Potential of Foie Gras." *Proceedings of the National Academy of Sciences* 104 (26): 10998–1.

Sponheimer, M., and J. A. Lee-Thorp. 1999. "Isotopic Evidence for the Diet of an Early Hominid, Australopithecus Africanus." *Science* 283 (5400): 368–70.

Sterzenbach, Torsten, Ralf Helbig, Christian Hannig, and Matthias Hannig. 2020. "Bioadhesion in the Oral Cavity and Approaches for Biofilm Management by Surface Modifications." *Clinical Oral Investigations* 24 (12): 4237–60.

Stewart, Nicolas Andre, Raquel Fernanda Gerlach, Rebecca L. Gowland, Kurt J. Gron, and Janet Montgomery. 2017. "Sex Determination of Human Remains from Peptides in Tooth Enamel." *Proceedings of the National Academy of Sciences* 114 (52): 13649–54.

Strixner, T., and U. Kulozik. 2011. "7 - Egg Proteins." In *Handbook of Food Proteins*, edited by G. O. Phillips and P. A. Williams, 150–209. Woodhead Publishing.

Suryanarayan, Akshyeta, Miriam Cubas, Oliver E. Craig, Carl P. Heron, Vasant S. Shinde, Ravindra N. Singh, Tamsin C. O'Connell, and Cameron A. Petrie. 2021. "Lipid Residues in Pottery from the Indus Civilisation in Northwest India." *Journal of Archaeological Science* 125: 105291.

Taylor, William Timothy Treal, Julia Clark, Jamsranjav Bayarsaikhan, Tumurbaatar Tuvshinjargal, Jessica Thompson Jobe, William Fitzhugh, Richard Kortum, et al. 2020. "Early Pastoral Economies and Herding Transitions in Eastern Eurasia." *Scientific Reports* 10 (1): 1001.

Teasdale, Matthew D., Sarah Fiddyment, Jiří Vnouček, Valeria Mattiangeli, Camilla Speller, Annelise Binois, Martin Carver, et al. 2017. "The York Gospels: A 1000-Year Biological Palimpsest." *Royal Society Open Science* 4 (10): 170988.

Tieszen, Larry L., and Tim Fagre. 1993. "Effect of Diet Quality and Composition on the Isotopic Composition of Respiratory CO2, Bone Collagen, Bioapatite, and Soft Tissues." In *Prehistoric Human Bone: Archaeology at the Molecular Level*, edited by Joseph B. Lambert and Gisela Grupe, 121–55. Berlin, Heidelberg: Springer Berlin Heidelberg.

Tromp, Monica, and John V. Dudgeon. 2015. "Differentiating Dietary and Non-Dietary Microfossils Extracted from Human Dental Calculus: The Importance of Sweet Potato to Ancient Diet on Rapa Nui." *Journal of Archaeological Science* 54: 54–63.

Tsolis, Antonios C., Nikos C. Papandreou, Vassiliki A. Iconomidou, and Stavros J. Hamodrakas. 2013. "A Consensus Method for the Prediction of 'Aggregation-Prone' Peptides in Globular Proteins." *PloS One* 8 (1): e54175.

Tsutaya, Takumi, Meaghan Mackie, Rikai Sawafuji, Takako Miyabe-Nishiwaki, Jesper V. Olsen, and Enrico Cappellini. 2021. "Faecal Proteomics as a Novel Method to Study Mammalian Behaviour and Physiology." *Molecular Ecology Resources* 21 (6): 1808–19.

UniProt Consortium. 2021. "UniProt: The Universal Protein Knowledgebase in 2021." *Nucleic Acids Research* 49 (D1): D480–89.

Van der Lee, Robin, Marija Buljan, Benjamin Lang, Robert J. Weatheritt, Gary W. Daughdrill, A. Keith Dunker, Monika Fuxreiter, et al. 2014. "Classification of Intrinsically Disordered Regions and Proteins." *Chemical Reviews* 114 (13): 6589–6631.

Vanderwarker, A., and T. M. Peres. 2010. *Integrating Zooarchaeology and Paleoethnobotany: A Consideration of Issues, Methods, and Cases*. Edited by Amber M. Van Der warker and Tanya M. Peres. Springer, New York, NY.

Vickery, B. P., S. Chin, and A. W. Burks. 2011. "Pathophysiology of Food Allergy." *Pediatric Clinics of North America*. 58(2): 363–376. doi:10.1016/j.pcl.2011.02.012.

Vigne, Jean-Denis, and Daniel Helmer. 2007. "Was Milk a 'secondary Product' in the Old World Neolithisation Process? Its Role in the Domestication of Cattle, Sheep and Goats." *Anthropozoologica* 42 (2): 9–40.

Vogel, J. C., and Nikolaas J. van der Merwe. 1977. "Isotopic Evidence for Early Maize Cultivation in New York State." *American Antiquity* 42 (2): 238–42.

Wagner, A., K. K. Richter, E. Ludes, and R. M. Arbogast. 2020. "Whale Bone Puzzles: Reconstructing and Identifying Historical Whale Skeletons Using Archive Records, Osteology, and Zooarchaeology by Mass Spectrometry (ZooMS)." *Journal of Conservation and Museum Studies* 18 (1): 1–12.

Wang, Chuan-Chao, Sabine Reinhold, Alexey Kalmykov, Antje Wissgott, Guido Brandt, Choongwon Jeong, Olivia Cheronet, et al. 2019. "Ancient Human Genome-Wide Data from a 3000-Year Interval in the Caucasus Corresponds with Eco-Geographic Regions." *Nature Communications* 10 (1): 590.

Wang, D., X. Zhou, J. Deng, and C. Fang. 2021. "Preparation and Characterization of Crystalline Hydroxyapatite Induced by Self-Assembled Peptide and the Potential Application in Remineralizing Dentin." *Advanced Engineering Materials*. 23: 2001470. https://doi.org/10.1002/adem.202001470

Warinner, C., J. Hendy, C. Speller, E. Cappellini, R. Fischer, C. Trachsel, J. Arneborg, et al. 2014. "Direct Evidence of Milk Consumption from Ancient Human Dental Calculus." *Scientific Reports* 4: 7104.

Warinner, Christina, João F. Matias Rodrigues, Rounak Vyas, Christian Trachsel, Natallia Shved, Jonas Grossmann, Anita Radini, et al. 2014. "Pathogens and Host Immunity in the Ancient Human Oral Cavity." *Nature Genetics* 46 (4): 336–44.

Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer, et al. 2018. "SWISS-MODEL: Homology Modelling of Protein Structures and Complexes." *Nucleic Acids Research* 46 (W1): W296–303.

Welker, Frido. 2018. "Palaeoproteomics for Human Evolution Studies." *Quaternary Science Reviews* 190: 137–47.

Welker, Frido, Matthew J. Collins, Jessica A. Thomas, Marc Wadsley, Selina Brace, Enrico Cappellini, Samuel T. Turvey, et al. 2015. "Ancient Proteins Resolve the Evolutionary History of Darwin's South American Ungulates." *Nature* 522 (7554): 81–84.

Welker, Frido, Jazmín Ramos-Madrigal, Petra Gutenbrunner, Meaghan Mackie, Shivani Tiwary, Rosa Rakownikow Jersie-Christensen, Cristina Chiva, et al. 2020. "The Dental Proteome of Homo Antecessor." *Nature* 580 (7802): 235–38.

Wellman-Labadie, O., R. Lakshminarayanan, and M. Hencke. 2008. "Antimicrobial Properties of Avian Eggshell-specific C-type Lectin-like Proteins." *FEBS Letters* 582 (5): 699–704.

Weyrich, Laura S., Sebastian Duchene, Julien Soubrier, Luis Arriola, Bastien Llamas, James Breen, Alan G. Morris, et al. 2017. "Neanderthal Behaviour, Diet, and Disease Inferred from Ancient DNA in Dental Calculus." *Nature* 544 (7650): 357–61.

Wilkin, Shevan, Alicia Ventresca Miller, William T. T. Taylor, Bryan K. Miller, Richard W. Hagan, Madeleine Bleasdale, Ashley Scott, et al. 2020. "Dairy Pastoralism Sustained Eastern Eurasian Steppe Populations for 5,000 Years." *Nature Ecology & Evolution* 4 (3): 346–55.

Wilkinson, Keith, Chris Stevens, and Elizabeth Sidell. 2003. *Environmental Archaeology: Approaches, Techniques & Applications*. Tempus.

Wiśniewski, Jacek R., Alexandre Zougman, Nagarjuna Nagaraj, and Matthias Mann. 2009. "Universal Sample Preparation Method for Proteome Analysis." *Nature Methods* 6 (5): 359–62.

Wolverton, Steve. 2013. "Data Quality in Zooarchaeological Faunal Identification." *Journal of Archaeological Method and Theory* 20 (3): 381–96.

Wright, Patti J. 2010. "Methodological Issues in Paleoethnobotany: A Consideration of Issues, Methods, and Cases." In *Integrating Zooarchaeology and Paleoethnobotany: A Consideration of Issues, Methods, and Cases*, edited by Amber M. Van Der warker and Tanya M. Peres, 37–64. New York, NY: Springer New York.

Xie, Mingsi, Anna Shevchenko, Binghua Wang, Andrej Shevchenko, Changsui Wang, and Yimin Yang. 2016. "Identification of a Dairy Product in the Grass Woven Basket from Gumugou Cemetery (3800 BP, Northwestern China)." *Quaternary International: The Journal of the International Union for Quaternary Research* 426: 158–65.

Xue, Bin, Daniel Soeria-Atmadja, Mats G. Gustafsson, Ulf Hammerling, A. Keith Dunker, and Vladimir N. Uversky. 2011. "Abundance and Functional Roles of Intrinsic Disorder in Allergenic Proteins and Allergen Representative Peptides." *Proteins* 79 (9): 2595–2606.

Yang, Yimin, Anna Shevchenko, Andrea Knaust, Idelisi Abuduresule, Wenying Li, Xingjun Hu, Changsui Wang, and Andrej Shevchenko. 2014. "Proteomics Evidence for Kefir Dairy in Early Bronze Age China." *Journal of Archaeological Science* 45 (May): 178–86.

Zelensky, Alex N., and Jill E. Gready. 2005. "The C-Type Lectin-like Domain Superfamily." *The FEBS Journal* 272 (24): 6179–6217.

Zhang, Jing, Jian Wang, Chengwei Ma, and Junxia Lu. 2020. "Hydroxyapatite Formation Coexists with Amyloid-like Self-Assembly of Human Amelogenin." *International Journal of Molecular Sciences* 21 (8). https://doi.org/10.3390/ijms21082946.

Zischler, H., M. Höss, O. Handt, A. von Haeseler, A. C. van der Kuyl, and J. Goudsmit. 1995. "Detecting Dinosaur DNA." *Science*.

# 11. Summary

This dissertation presents research that was produced through proteomic analysis of ancient dental calculus or biomineralized plaque. The potential of this substrate lies in its ability to trap and preserve biomolecules onto the surface of the teeth of ancient humans, providing a direct line of evidence for the consumption of particular dietary items. The goals of this thesis include recovering and identifying dietary food proteins as well as gaining a deeper understanding of the factors that affect the potential preservation and recovery of ancient proteins from dental calculus. While dairy proteins have been frequently recovered from ancient dental calculus, non-dairy dietary protein recovery has been extremely rare. A better understanding of the factors that influence the incorporation, recovery, and identification of ancient proteins from dental calculus is necessary to increase the potential of paleoproteomic studies of ancient diet.

In Manuscript A, we analyzed human calculus from individuals who died in the Kilkenny Union Workhouse during the Irish Famine through microremain and proteomic analyses. The protein analysis resulted from the research done for this thesis, while other researchers performed the microremain analysis. Some historical accounts of the available food supplies were available, and we compared our results to these records to determine how well our results could corroborate or add to written sources about foods available to individuals who lived in the workhouse and relied completely on the relief food provided to them. Our results provided evidence for the consumption of various foods, including oats, potatoes, corn, wheat, eggs, and milk. Microparticles specific to plants such as corn, barley, wheat, oats, and potato were recovered along with the milk protein β-lactoglobulin and the chicken egg protein ovalbumin. These results were consistent with historical accounts. Although chicken eggs were not often mentioned in the historical accounts of available foodstuff, they likely reflect a pre-famine dietary item, as calculus accumulates throughout an individual's life.

In Manuscript B, we analyzed the remains of the earliest inhabitants of Xinjiang discovered to date, dating from 3000 to 1700 BC. We presented both genomic and proteomic results. The protein analysis resulted from the research done for this thesis, while other researchers performed the genomic component. Our genomic data includes a total of 20 individuals from the Dzungarian Basin (ca. 3000-2800 BCE) and the Tarim Basin (ca. 2100-1700 BC). Dzungarian Basin individuals showed ancestry consistent with Afanasievo with an additional local contribution, while Tarim Basin individuals have only local ancestry. Our proteomic data of dental calculus from seven individuals from the earliest strata at the Xiaohe site in the Tarim Basin, show a strong signature of milk consumption from all seven individuals, which suggests that this group relied on dairy technology from its founding. Our results failed to support previous hypotheses about the origins of the Tarim Basin mummies. Instead, they suggest that the earliest Tarim Basin cultures emerged from a genetically isolated local population that had already adopted neighboring pastoralist technologies, despite their genetic isolation.

In Manuscript C, we used proteomic methods to investigate the subsistence basis of individuals buried in the North Caucasus and surrounding regions of the South Caucasus, Oka-Don Volga, and Ural Mountains from the Eneolithic through the Roman era, by identifying dairy proteins recovered from their dental calculus. Previous archaeological and archaeogenetic evidence points to the Pontic-Caspian steppe zone as the region from which the earliest mobile pastoralists merged and spread, eventually influencing populations across Eurasia. However, little is known about their economic bases and the factors that contributed to their heightened mobility. Our protein results demonstrate that sheep milking occurred

with the earliest forms of Eneolithic pastoralism in the North Caucasus. Later, during the 4th millennium BC, Maykop and early Yamnaya populations also focused exclusively on sheep milk. We observe an economic diversification of dairy herds that coincide with aridification during the subsequent Late Yamnaya and North Caucasus cultural phases, followed by severe climate deterioration during the Catacomb and Lola periods. The high mobility associated with the Middle and Late Bronze Ages may have resulted from a need for additional pasture to feed herds. Later, during the Iron Age, the North Caucasian steppe was repopulated, and we find evidence of dairying of sheep, goats, and cattle as well as the introduction of horse dairying.

In Manuscript D, we examined the dental calculus of individuals buried at the southern Levant Bronze and Iron Age sites of Megiddo and Tel Erani using proteomic and microremain techniques- the first study of its kind in the ancient Near East. The protein analysis resulted from the research done for this thesis, while other researchers performed the microremain analysis. We confirmed the presence of staple foods such as wheat, millet, and date palm. In addition, we found clear evidence for sesame seed proteins, confirming that sesame had become an essential oil-bearing plant in the Levant by the 2nd millennium BC. We also found the earliest direct evidence for the spice turmeric, which predates the earliest written record for this spice in the region by approximately 900 years. We have also found evidence for the consumption of soy proteins. We recovered credible evidence for banana in the Iron Age Levant, which provides a critical geographic link between the plant's origins in New Guinea (5th millennium BC) and its appearance in West Africa (Cameroon) in the 1st century BC.

Finally, this dissertation examined the patterns of protein recovery and identification from dental calculus extraction and other ancient samples and presents a series of observations and hypotheses on the possible mechanisms of protein preservation in archaeological contexts.

## 12. Zusammenfassung

In dieser Dissertation werden Forschungsergebnisse vorgestellt, die durch die proteomische Analyse von altem Zahnstein oder biomineralisiertem Zahnbelag gewonnen wurden. Das Potenzial dieses Substrats liegt in seiner Fähigkeit, Biomoleküle auf der Oberfläche der Zähne alter Menschen zu binden und zu konservieren, wodurch ein direkter Nachweis für den Verzehr bestimmter Nahrungsbestandteile erbracht werden kann. Zu den Zielen dieser Arbeit gehören die Gewinnung und Identifizierung von Nahrungsproteinen sowie die Erlangung eines tieferen Verständnisses der Faktoren, die die potenzielle Konservierung und Gewinnung alter Proteine aus Zahnstein beeinflussen. Während Milchproteine häufig aus altem Zahnstein geborgen wurden, ist die Gewinnung von Nicht-Milchproteinen aus Nahrungsmitteln extrem selten gewesen. Ein besseres Verständnis der Faktoren, die den Einbau, die Wiedergewinnung und die Identifizierung alter Proteine aus Zahnstein beeinflussen, ist notwendig, um das Potenzial von paläoproteomischen Studien über alte Ernährung zu erhöhen.

In Manuskript A wurde menschlicher Zahnstein von Personen, die während der irischen Hungersnot im Kilkenny Union Workhouse starben, mittels Mikrorest- und Proteomanalysen untersucht. Die Proteinanalyse wurde im Rahmen dieser Arbeit durchgeführt, während andere Forscher die Mikrospurenanalyse durchführten. Wir haben unsere Ergebnisse mit diesen Aufzeichnungen verglichen, um festzustellen, inwieweit unsere Ergebnisse die schriftlichen Quellen über die Nahrungsmittel, die den Menschen zur Verfügung standen, die im Arbeitshaus lebten und sich vollständig auf die ihnen zur Verfügung gestellte Hilfsnahrung verließen, bestätigen oder ergänzen konnten. Unsere Ergebnisse lieferten Beweise für den Verzehr verschiedener Lebensmittel, darunter Hafer, Kartoffeln, Mais, Weizen, Eier und Milch. Mikropartikel, die für Pflanzen wie Mais, Gerste, Weizen, Hafer und Kartoffeln spezifisch sind, wurden ebenso gefunden wie das Milchprotein β-Lactoglobulin und das Hühnerei-Protein Ovalbumin. Diese Ergebnisse stimmen mit historischen Berichten überein. Obwohl Hühnereier in den historischen Berichten über die verfügbaren Nahrungsmittel nicht oft erwähnt wurden, spiegeln sie wahrscheinlich einen Nahrungsbestandteil vor der Hungersnot wider, da sich Kalk im Laufe des Lebens eines Menschen ansammelt.

In Manuskript B analysierten wir die Überreste der frühesten bisher entdeckten Bewohner von Xinjiang, die aus der Zeit zwischen 3000 und 1700 v. Chr. stammen. Wir präsentierten sowohl genomische als auch proteomische Ergebnisse. Die Proteinanalyse wurde im Rahmen dieser Arbeit durchgeführt, während der genomische Teil von anderen Forschern übernommen wurde. Unsere genomischen Daten umfassen insgesamt 20 Individuen aus dem Dzungarischen Becken (ca. 3000-2800 v. Chr.) und dem Tarimbecken (ca. 2100-1700 v. Chr.). Die Individuen aus dem Dzungarischen Becken zeigten eine Abstammung, die mit der von Afanasievo übereinstimmt, mit einem zusätzlichen lokalen Beitrag, während die Individuen aus dem Tarimbecken nur eine lokale Abstammung haben. Unsere proteomischen Daten des Zahnsteins von sieben Individuen aus den frühesten Schichten der Xiaohe-Fundstelle im Tarim-Becken zeigen bei allen sieben Individuen eine deutliche Signatur von Milchkonsum, was darauf hindeutet, dass diese Gruppe seit ihrer Gründung auf die Milchwirtschaft angewiesen war. Unsere Ergebnisse konnten frühere Hypothesen über die Herkunft der Mumien aus dem Tarimbecken nicht bestätigen. Stattdessen deuten sie darauf hin, dass die frühesten Kulturen des Tarimbeckens aus einer genetisch isolierten lokalen Bevölkerung hervorgingen, die trotz ihrer genetischen Isolation bereits Technologien der benachbarten Hirtenvölker übernommen hatte.

In Manuskript C untersuchten wir mit proteomischen Methoden die Lebensgrundlage von

Individuen, die im Nordkaukasus und den umliegenden Regionen des Südkaukasus, der Oka-Don-Wolga und des Uralgebirges vom Eneolithikum bis in die Römerzeit bestattet wurden, indem wir Milchproteine aus ihrem Zahnstein identifizierten. Frühere archäologische und archäogenetische Beweise deuten darauf hin, dass die pontisch-kaspische Steppenzone die Region ist, in der sich die ersten mobilen Hirtenvölker zusammenschlossen und ausbreiteten und schließlich die Bevölkerungen in ganz Eurasien beeinflussten. Über ihre wirtschaftlichen Grundlagen und die Faktoren, die zu ihrer erhöhten Mobilität beitrugen, ist jedoch wenig bekannt. Unsere Proteinergebnisse zeigen, dass das Melken von Schafen schon bei den frühesten Formen des eneolithischen Pastoralismus im Nordkaukasus vorkam. Später, im 4. Jahrtausend v. Chr., konzentrierten sich auch die Maykop- und frühen Yamnaya-Bevölkerungen ausschließlich auf Schafsmilch. Wir beobachten eine wirtschaftliche Diversifizierung der Milchviehherden, die mit der Aridifizierung während der nachfolgenden späten Yamnaya- und Nordkaukasus-Kulturphasen zusammenfällt, gefolgt von einer schweren Klimaverschlechterung während der Katakomben- und Lola-Periode. Die hohe Mobilität in der mittleren und späten Bronzezeit könnte auf den Bedarf an zusätzlichem Weideland zur Fütterung der Herden zurückzuführen sein. Später, in der Eisenzeit, wurde die nordkaukasische Steppe wieder besiedelt, und wir finden Belege für die Haltung von Schafen, Ziegen und Rindern sowie für die Einführung der Pferdemast.

In Manuskript D untersuchten wir den Zahnstein von Individuen, die an den bronze- und eisenzeitlichen Fundstätten Megiddo und Tel Erani in der südlichen Levante bestattet wurden, mit Hilfe von Proteom- und Mikrorest-Techniken - die erste Studie dieser Art im alten Nahen Osten. Die Proteinanalyse wurde im Rahmen dieser Arbeit durchgeführt, während andere Forscher die Mikrorestanalyse durchführten. Wir konnten das Vorhandensein von Grundnahrungsmitteln wie Weizen, Hirse und Dattelpalme bestätigen. Außerdem fanden wir eindeutige Beweise für Sesamsamenproteine, die bestätigen, dass Sesam in der Levante im 2. Jahrtausend v. Chr. zu einer wesentlichen Ölpflanze geworden war. Wir haben auch die frühesten direkten Belege für das Gewürz Kurkuma gefunden, die den frühesten schriftlichen Aufzeichnungen über dieses Gewürz in der Region um etwa 900 Jahre vorausgehen. Wir haben auch Belege für den Verzehr von Sojaproteinen gefunden. Wir haben glaubwürdige Belege für die Banane in der eisenzeitlichen Levante gefunden, die eine wichtige geografische Verbindung zwischen den Ursprüngen der Pflanze in Neuguinea (5. Jahrtausend v. Chr.) und ihrem Auftauchen in Westafrika (Kamerun) im ersten Jahrhundert v. Chr. herstellen.

Schließlich wurden in dieser Dissertation die Muster der Proteingewinnung und -identifizierung aus Zahnstein und anderen antiken Proben untersucht und eine Reihe von Beobachtungen und Hypothesen zu den möglichen Mechanismen der Proteinkonservierung in archäologischen Kontexten vorgestellt.

# 13. Declaration of Honour (Eigenständigkeitserklaerung)

Ashley Scott

Max Planck Institute for the Science of Human History, Jena Germany
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

In accordance with the doctoral degree regulations of the Faculty of Biological Sciences at the Friedrich Schiller University Jena, I hereby declare:

(a) that I am aware of the applicable doctoral regulations,

(b) that I have written the doctoral thesis myself and that I have not taken any text sections from another author or from my own examination papers without indicating them and that I have indicated all tools and sources used by myself in this work,

(c) that I have mentioned all persons who have supported me in the selection and evaluation of the material as well as in the production of the manuscript,

(d) that I have not used the assistance of a commercial doctoral advisor and that third parties have neither directly nor indirectly received monetary benefits from me for work that is related to the content of the submitted doctoral thesis,

(e) that I have not yet submitted this doctoral thesis as an examination paper for an academic examination,

(f) that I have submitted neither the same thesis nor an essentially similar thesis, nor a different thesis as a doctoral thesis at another university.

# 14. Acknowledgements

First, I would like to thank my advisor and mentor, Dr. Christina Warinner, for taking a chance on me and allowing me to be a part of her research group in the fascinating and emerging field of paleoproteomics as a part of the Max Planck Society at the Max Planck Institutes for the Science of Human History and Evolutionary Anthropology. Thank you for introducing me to so many incredible research projects and freely and generously sharing your knowledge and experience.

I would also like to thank Philipp Stockhammer for allowing me to work on his ERC project, FoodTransforms, and giving me access to an incredible collection of samples for my PhD research. Your intense enthusiasm for the work we do has always been inspiring.

Thank you to all of the PIs, group leaders, and post-docs that provided me with feedback and knowledge during the course of my Ph.D. Thank you to all of my collaborators for your hard work, patience, and encouragement as I learned the processes of manuscript construction and collaborative science. Thank you to all of my fellow students, who kept me sane and laughing throughout the last few years. To my lovely office mates, thank you for the dog-sitting, the movies, the emotional support, and for being so wonderful and for also sharing your knowledge and experience so freely. I would also like to thank all of the Max Planck staff for the hard work that goes into running an Institute efficiently. I would specifically like to thank Anke Trinkler, who went above and beyond the role of administrative assistant and helped me navigate the complex and terrifying world of international travel and study in a country where I did not speak the language or understand the bureaucracy. I could not have done it without you.

I'd like to thank and acknowledge the McNair Scholars program and Diana Elrod, in particular, for helping me navigate the college and graduate system as a first-generation college student. There were many times when I was in over my head, confused and lost, and after a meeting with Dr. Elrod, everything would be better. You were an incredible advocate, and I am infinitely grateful for your formative role in shaping my early research career.

I also want to thank Barney Venables and Steve Wolverton, who took a chance on me, gave me a new start during my Master's studies, and introduced me to the wonderful world of paleoproteomics. The technical skills I learned from Dr. Veneables have been immeasurably helpful in my journey. Dr. Wolverton's kindness, encouragement, and advice led me to where I am now. Before I joined their labs, I was on the verge of leaving my Master's program. The support from both of these mentors (and feminists) kept me from leaving science behind, and the decision to join their labs was one of the best I ever made.

Finally, I would like to thank all of my family for their continuous support, encouragement, and love. I want to thank my father for encouraging my curiosity and love of learning at an early age and my mother, who always found my extremely stubborn nature an endless source of amusement and not something to be curbed. It has served me well during this Ph.D. My siblings, who are my best friends, you managed to keep me laughing with stupid memes and dark humor, even through the most challenging parts of my academic journey. And finally, my husband Brent (my most favorite person in the world). Thank you for loving me and for all of your sacrifices over the last several years. You are my rock, and I couldn't have done it without you.

# 15. Appendix

**Manuscript Nr.:** 1 (Manuscript A)
**Short reference:** Geber J. et al. (2019), *PNAS*
**Contribution of the doctoral student:** JG and MT conceived and designed this research study with input from CW AS and JH. AS provided the analysis of the proteomic portion of the study.
**Contribution of the doctoral student to figures reflecting experimental data (only for original articles):** Proteomic figures in the Supplemental data were generated from the analysis provided by AS. JG, MT, and AS composed the manuscript with contributions from JH and CW. Figure 1 in the manuscript was an historical illustration provided courtesy of the National Library of Ireland. Figure 2 of the manuscript depicted data generated and analyzed by MT.

____________________ ____________________________________

Unterschrift Kandidat/-in Unterschrift Betreuer/-in (Mitglied der Fakultät)

**Manuscript Nr.:** 2 (Manuscript B)
**Short reference :** Zhang et al. (2021), *Nature*
**Contribution of the doctoral student:** AS contributed the proteomic component of the study, with feedback from CW.
**Contribution of the doctoral student to figures reflecting experimental data (only for original articles):**
The final three figures in the manuscript represented the genomic data generated for the manuscript, and therefore, AS did not contribute to these figures. Extended Figure 5. was generated using data from the analysis provided by AS.

____________________ ____________________________________

Unterschrift Kandidat/-in Unterschrift Betreuer/-in (Mitglied der Fakultät)

**Manuscript Nr.:** 3 (Manuscript C)
**Short reference :** Scott. et al. (2021), in submission Nature Evolution and Ecology
**Contribution of the doctoral student:** AS performed 90% of the experimental laboratory work, and 100% of the data analysis for the manuscript, with feedback from CW. AS wrote the original draft with feedback from SR and CW. AS, CW, SR, TH, and WH refined the manuscript.
**Contribution of the doctoral student to figures reflecting experimental data (only for original articles):** Figures 1, 2, and 3 of the manuscript were generated from original data produced by AS. AS and CW both designed the figures. CW created the illustrations with help and feedback from AS.

______________________ ________________________________________

Unterschrift Kandidat/-in Unterschrift Betreuer/-in (Mitglied der Fakultät)

**Manuscript Nr.:** 4 (Manuscript D)
**Short reference :** Scott, A. et al. (2021), PNAS
**Contribution of the doctoral student:** AS performed 100% of the experimental laboratory work, and 100% of the data analysis for the proteomic portion of the manuscript, with feedback from CW, while RP performed 100% of the macroremain analysis. AS wrote the original draft with feedback from PS and CW. AS, CW, PS, VAW, and RP refined the manuscript.
**Contribution of the doctoral student to figures reflecting experimental data (only for original articles):** Figure 1 was designed and illustrated by AS using data generated from previous archaeological identifications of specific dietary items across Eurasia. AS collected the data, coordinates, and original citations used for the figure. Figure 2 was provided by a co-author. Figure 3 was made using the data generated from co-author RP. Figure 4 was made using original data generated by AS. AS and CW designed the figure. CW illustrated the figure with feedback from AS.

______________________ ________________________________________

Unterschrift Kandidat/-in Unterschrift Betreuer/-in (Mitglied der Fakultät)