

Research



Cite this article: Runge AKW *et al.* 2021 Palaeoproteomic analyses of dog palaeofaeces reveal a preserved dietary and host digestive proteome. *Proc. R. Soc. B* **288**: 20210020. <https://doi.org/10.1098/rspb.2021.0020>

Received: 7 January 2021

Accepted: 10 June 2021

Subject Category:

Palaeobiology

Subject Areas:

biochemistry, behaviour, evolution

Keywords:

dogs, palaeofaeces, palaeoproteomics, zooms, Nunalleq Alaska, archaeology

Authors for correspondence:

Anne Kathrine W. Runge

e-mail: ak@palaeome.org

Camilla Speller

e-mail: camilla.speller@york.ac.uk

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5470468>.

Palaeoproteomic analyses of dog palaeofaeces reveal a preserved dietary and host digestive proteome

Anne Kathrine W. Runge^{1,2}, Jessica Hendy^{1,3}, Kristine K. Richter^{4,5}, Edouard Masson-MacLean⁶, Kate Britton^{6,7}, Meaghan Mackie^{2,8}, Krista McGrath^{1,9}, Matthew Collins^{2,10}, Enrico Cappellini² and Camilla Speller^{1,11}

¹BioArCh, Department of Archaeology, University of York, Environment Building, Wentworth Way, YO10 5DD York, UK

²Section for Evolutionary Genomics, the GLOBE Institute, University of Copenhagen, Øster Farimagsgade 5A, 1353 København K, Denmark

³Department of Archaeogenetics, and ⁴Department of Archaeology, Max Planck Institute for the Science of Human History, Kahlaische Strasse 10, 07743 Jena, Germany

⁵Department of Anthropology, Harvard University, Cambridge, MA 02138, USA

⁶Department of Archaeology, University of Aberdeen, Aberdeen, Scotland, UK

⁷Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig 04103

⁸The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3b, 2200 København N, Denmark

⁹Department of Prehistory and Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

¹⁰Department of Archaeology, University of Cambridge, Cambridge CB2 3DZ, UK

¹¹Department of Anthropology, University of British Columbia, 6303 NW Marine Drive, Vancouver, Canada V6T 1Z1

AKWR, 0000-0003-2421-4831; KKR, 0000-0003-3591-6900; MC, 0000-0003-4226-5501; CS, 0000-0001-7128-9903

The domestic dog has inhabited the anthropogenic niche for at least 15 000 years, but despite their impact on human strategies, the lives of dogs and their interactions with humans have only recently become a subject of interest to archaeologists. In the Arctic, dogs rely exclusively on humans for food during the winter, and while stable isotope analyses have revealed dietary similarities at some sites, deciphering the details of provisioning strategies have been challenging. In this study, we apply zooarchaeology by mass spectrometry (ZooMS) and liquid chromatography tandem mass spectrometry to dog palaeofaeces to investigate protein preservation in this highly degradable material and obtain information about the diet of domestic dogs at the Nunalleq site, Alaska. We identify a suite of digestive and metabolic proteins from the host species, demonstrating the utility of this material as a novel and viable substrate for the recovery of gastrointestinal proteomes. The recovered proteins revealed that the Nunalleq dogs consumed a range of Pacific salmon species (coho, chum, chinook and sockeye) and that the consumed tissues derived from muscle and bone tissues as well as roe and guts. Overall, the study demonstrated the viability of permafrost-preserved palaeofaeces as a unique source of host and dietary proteomes.

1. Introduction

The domestic dog (*Canis lupus familiaris*) originated in Eurasia [1,2] at least 15 000 years ago [3], but despite our extended collaborative history, the lives of dogs and their management through time has largely been neglected by

researchers [4]. However, attitudes towards the lived experiences of dogs and their roles in past human societies are beginning to shift [5], and in recent years, dietary studies have provided novel insight into their lifeways and how different cultures incorporated dogs into their subsistence strategies. Dog and human diets can be so similar that they can be analysed in lieu of humans as a proxy for reconstructing human diet [6] and have revealed a reliance on similar resources [7] although comparative studies indicate that this is not always the case [8]. Genetic analyses have further shown that as human subsistence practices changed with the spread of agriculture, the diet of non-arctic dogs was impacted as well [9].

The earliest known dog remains in North America indicate that this domesticated species was introduced at least 10 000 years ago and belonged to a now extinct lineage of arctic dog [10,11]. In the North American Arctic, dogs do not appear to have been present in large numbers until around 1000 BP [12]. The ancestors of Yup'ik, Inuit and other Indigenous Arctic groups relied on advanced transportation technology including dog traction [13] and appear to have been the first to introduce specialized sled dogs to North America [14]. Arctic dogs rely exclusively on humans for food during the long winters, but may have been fed differently or less frequently in summer, or let loose to fend for themselves [15,16]. Working sled dogs are a particularly expensive resource, requiring up to 3.2 kg of fish or meat every day [17] and provisioning of dogs would therefore have played a significant role in the food procurement strategies of past arctic cultures. Stable isotope analyses from coastal sites in Alaska [7,18,19] and Canada [20–22] have shown that the majority of dog diet in ancient and historical North American Arctic cultures consisted of marine mammals and salmonids. These studies provide important insight into the diet of dogs and humans, revealing similarities in resource allocation, but do not provide information on seasonal variation unless incrementally developed tissues such as hair or claw are targeted [7]. In addition, the practice of feeding specific species, such as walrus [17] or chum salmon (also known as 'dog salmon') [7], or animal parts, including hide, that were considered unpalatable or unsuitable for human consumption can be difficult to characterize in any detail using stable isotope approaches.

Recent advances in the field of palaeoproteomics have opened up new avenues for dietary studies of past populations, complementing more traditional zooarchaeology, palaeobotany and stable isotope approaches. Proteins are frequently tissue specific, allowing distinct plant parts (e.g. seeds, leaves and roots) and animal products (e.g. muscle, milk and blood) to be distinguished [23]. In particular, ancient dental calculus has emerged as a valuable substrate for preserving dietary biomolecules, revealing novel insights especially in ancient dairy consumption [24,25]. Palaeoproteomics has also been used to identify the 'last meal' of Ötzi, a human Copper Age mummy [26] and to reveal evidence of breast milk feeding from the rib bone of a neonatal dog [27], as well as food composition in a range of well-preserved archaeological materials [28–30]. Staining techniques have previously demonstrated the survival of proteins in palaeofaeces [31], a material which has long been used for macro- and microscopic dietary analyses [32]. However, palaeoproteomic analyses have only been applied to the substrate with limited success [33]. Ancient DNA evidence has shown palaeofaeces to preserve dietary

information as well as evidence of the gut microbiome [34,35], but the scope of these studies has been limited owing to its complex and highly degradable nature [32]. Furthermore, DNA studies are limited by a lack of tissue specificity, which is a major strength of palaeoproteomics. Here, in order to investigate the preservation of ancient proteins in palaeofaeces, and to reveal new insights into the subsistence of arctic dogs, we use shotgun palaeoproteomics to identify a range of host and dietary-derived proteins preserved in palaeofaeces and complement this analysis with zooarchaeology by mass spectrometry (ZooMS) on bone fragments collected from within this substrate.

2. Material and methods

(a) Samples

All samples were excavated from the Nunalleq site, southwest Alaska (figure 1) during research excavations in 2013–2015. The site, a pre-contact Yup'ik village close to the contemporary community at Quinagak, Alaska, was occupied between *ca* 1300 CE and 1750 CE. It has exceptional preservation of archaeological artefacts and bioarchaeological remains [36], including dog harnesses made of braided grass, dog bones, fur and even lice, as well as palaeofaeces [16]. Detailed site and sample information can be found in the electronic supplementary material, file S1.

(b) Palaeoproteomics

Five samples (figure 1; electronic supplementary material, table S1) were selected for protein analysis. Palaeoproteomic extraction was carried out following the protocol of Tsutaya *et al.* [37] designed for extraction of modern faeces. Full methodological details on protein extraction and data analysis can be found in the electronic supplementary material, file S1. Samples were analysed using liquid chromatography tandem mass spectrometry (LC-MS/MS) and spectral data were converted to MASCOT generic format (.mgf). Semi-tryptic MS/MS ion database searching was performed on MASCOT (Matrix Science™, v. 2.6.1) against a database composed of Swiss-Prot, and the *Canis lupus familiaris* and *Oncorhynchus mykiss* proteomes. Only proteins with a minimum of two non-identical peptides and those retained after a false discovery rate (FDR) estimation of 1% of distinct peptide spectrum matches were considered in further downstream interpretation. Keratins are considered separately (electronic supplementary material, table S3), while peptides identified in the extraction negative were excluded from further analysis along with matches to the reagents trypsin and lysyl peptidase (electronic supplementary material, table S4). Lowest common ancestor analysis was performed using Unipept metaproteomics analysis [38]. The taxonomy of peptides unclassified by Unipept were explored using protein–protein BLAST against all non-redundant protein sequences. STRING v. 11.0 [39] was used to analyse protein–protein interactions and proteome function using gene ontology (GO) annotations (electronic supplementary material, table S5), and expression profiles using human homologues was explored using the GENE CARDS suite [40].

(c) Zooarchaeology by mass spectrometry

The samples Nun-2 and Nun-4 each contained five bone fragments, which were analysed using ZooMS. Two pretreatment methods involving either acid demineralization (method A) or an ammonium bicarbonate buffer (method B) were used prior to collagen extraction. Samples were processed according to previously established protocols [41,42] and analysed on a MALDI-TOF mass spectrometer. MALDI spectra were processed using the R package MALDIquant [43] and visualized in mMass [44].

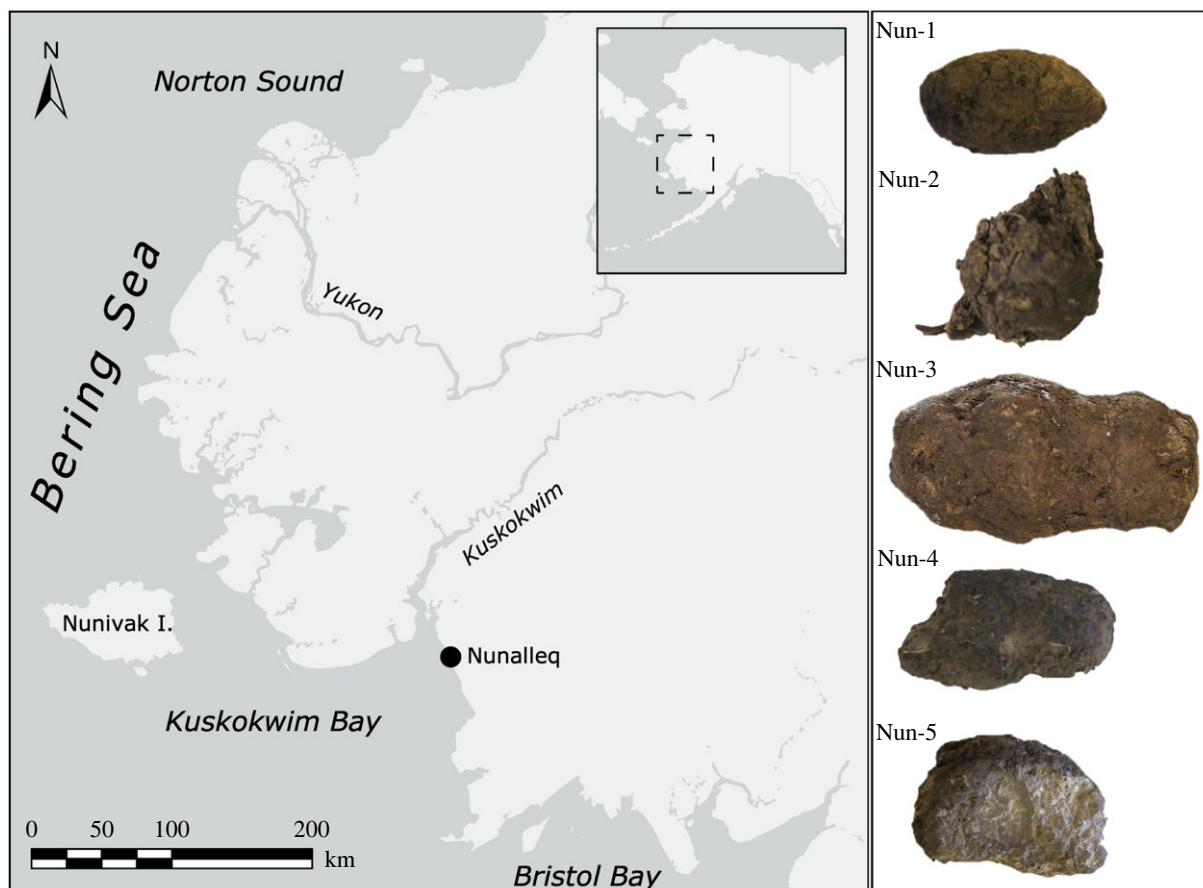


Figure 1. Map showing the location of the Nunalleq site in Alaska and pictures of the palaeofaeces analysed in this study.

Identifications (electronic supplementary material, table S6) were made based on published markers for mammals [41,45–48] (electronic supplementary material, table S7) and fishes including Pacific salmonids [49] (electronic supplementary material, tables S8 and S9). Full methodological details can be found in the electronic supplementary material.

3. Results and discussion

(a) Metaproteomics

We applied shotgun proteomic analysis to five palaeofaeces, revealing evidence of a digestive proteome as well as indicators of consumed food. After quality filtering, we detected a total of 83 proteins across all five samples, with 56 distinct individual proteins represented by 282 unique peptide sequences (excluding keratins) (electronic supplementary material, table S2; figure 2). Protein recovery varied between samples (figure 3), with Nun-3 showing evidence of 38 identified proteins and Nun-5 limited to two (electronic supplementary material, tables S1 and S2). All samples, except Nun-5, displayed proteins that could be classified as host-derived (table 1) and dietary (electronic supplementary material, table S10). In this study, the extraction method was selected to enhance the detection of dietary proteins by pelleting bacteria and other particulates present in the samples. As expected, we did not detect abundant microbial proteins related to the gut or faecal microbiome, although a few bacterial peptides were identified despite this methodological approach.

Four samples, Nun-1, Nun-2, Nun-3 and Nun-4, were excavated from house floor contexts and displayed superior protein preservation compared to Nun-5, which was excavated from a

debris context (electronic supplementary material, table S1). It is likely that the archaeological context influenced biomolecular preservation of the palaeofaeces, with house floors offering protection that debris contexts did not. However, given the absence of potential proteins in Nun-5 and considering that the sample did not contain hair or bone fragments, it is possible that it was misidentified as palaeofaeces. As a result, this sample is not considered in further analyses.

(b) Host proteins

(i) Host identification

First, we characterized the palaeofaecal proteome derived from the putative host. In accordance with the initial morphological characterization, we identified proteins associated with the gastrointestinal (GI) system, which were specific to the taxonomic orders of *Canis lupus familiaris* in all four successful samples. Further proteins were assigned to the clades Canidae, Caniformia and Carnivora. As some proteins are conserved across taxonomic units, proteins assigned to these clades were considered to have originated from dogs (table 1). Only a single protein assigned to Hominidae, neutrophil defensin 1, was identified in the dataset. Given the abundance of proteins derived from Carnivora, especially proteins associated with the digestive system, the combined evidence points to palaeofaeces deposited by canids.

(ii) Digestive proteins

LC-MS/MS analysis revealed evidence of a suite of proteins associated with the GI system (table 1), including digestion. These included, but were not limited to, colipase and inactive

Table 1. Proteins assigned to the taxonomic level of carnivora or lower. (The taxonomic assignment of the protein accession is based on the peptide with the most specific taxonomy detected within a sample. Gene names are those associated with the Uniprot accession. Expression data was obtained from human homologues in GeneCards.)

sample	identified host protein	Uniprot accession	gene name	most specific taxonomy	MAscOT score	expression
Nun-1	aminopeptidase N	AOA5F4CJ36 ^a	ANPEP	Canidae	84	PJ, SI
	colipase	P19090 ^a	CLPS	Caniformia	67	PJ
	deleted in malignant brain tumours 1	AOA5F4BZQ1	DMBT1	<i>Canis lupus familiaris</i>	76	SI, C, PJ
	dipeptidyl peptidase 4	F1PP08 ^a	DPP4	Canidae	97	MT incl. SI
	serpin family B member 6	E2RGP2	SERPINB6	<i>Canis lupus familiaris</i>	52	MT incl. PJ SI, C
	trefoil factor 2	F1PPU3 ^a	TFF2	Canidae	90	PJ, S
	zymogen granule protein 16	E2RS34	ZG16	Canidae	161	SI, C, R
	Nun-2	anionic trypsin	P06872 ^a	PRSS2	<i>Canis lupus familiaris</i>	98
colipase		P19090 ^a	CLPS	<i>Canis lupus familiaris</i>	69	PJ
peptidase S1 domain-containing protein		F1PCE8	LOC475521	Canidae	81	N/A
serpin family B member 6		E2RGP2	SERPINB6	<i>Canis lupus familiaris</i>	185	MT incl. PJ SI, C
syncollin		F1PW99	SYCN	<i>Canis lupus familiaris</i>	71	PJ, P
trefoil factor 2		F1PPU3 ^a	TFF2	Canidae	64	PJ, S
zymogen granule protein 16		E2RS34	ZG16	Canidae	103	SI, C, R
Nun-3		alpha 2-HS glycoprotein	E2QUV3	AHSG	Canidae	108
	chymotrypsin like	F1PA60	CTRL	Canidae	53	PJ, P
	colipase	P19090 ^a	CLPS	<i>Canis lupus familiaris</i>	345	PJ
	collagen alpha-2(I) chain	F1PHY1	COL1A2	Canidae	344	MT incl. SI, C, R
	dipeptidyl peptidase 4	F1PP08 ^a	DPP4	Canidae	66	MT incl. SI
	GLOBIN domain-containing protein	E2RLH6	LOC609402	Caniformia	193	MT incl. PJ, C, R
		J9JHF7	LOC100855540	Caninae	86	N/A
	inactive pancreatic lipase-related protein 1	AOA5F4CND7	PNLIP	<i>Canis lupus familiaris</i>	109	PJ
	peptidase S1 domain-containing protein	AOA5F4CBC4	CTRB2	<i>Canis lupus familiaris</i>	34	PJ, P, R, gut (fetal)
		AOA5F4CGL2	KLK1	<i>Canis lupus familiaris</i>	80	PJ, C, R
		AOA5F4DFJ6 ^a	PRSS2	<i>Canis lupus familiaris</i>	65	PJ
		F1PCE8	LOC475521	Canidae	89	N/A
		F6XMJ9	LOC478220	<i>Canis lupus familiaris</i>	91	N/A
		E2RGP2	SERPINB6	<i>Canis lupus familiaris</i>	117	MT incl. PJ SI, C
	serpin family B member 6	E2RGP2	SERPINB6	<i>Canis lupus familiaris</i>	117	MT incl. PJ SI, C
serpin family F member 1	AOA5F4CK29	SERPINF1	Canidae	86	MT incl. C, P, R, S	
serum albumin	P49064	ALB	Carnivora	102	MT incl. PJ, C, R	
trefoil factor 2	F1PPU3 ^a	TFF2	Canidae	70	PJ, S	
zymogen granule protein 16	E2RS34	ZG16	Canidae	214	SI, C, R	
Nun-4	colipase	P19090 ^a	CLPS	<i>Canis lupus familiaris</i>	49	PJ
	trefoil factor 2	F1PPU3 ^a	TFF2	Canidae	51	PJ, S
	zymogen granule protein 16	E2RS34	ZG16	Canidae	211	SI, C, R

^aDenotes protein accessions involved in digestive and metabolic functions based on GO annotations. Expression abbreviations explained: B, blood; C, colon; L, liver; MT, multiple tissues; P, pancreas; PJ, pancreatic juice; R, rectum; S, stomach; SI, small intestine.

(electronic supplementary material, figure S1). Although trefoil factor 2 is among the GO terms assigned to digestion (electronic supplementary material, table S5), the STRING

analysis (electronic supplementary material, figure S1) does not identify interactions with other host proteins identified here.

Digestion is facilitated by the secretion of fluids from the salivary glands, stomach, pancreas and small intestine [50]. Using data from UniProtKB and GeneCards, we also explored the tissues from which the identified host proteins were expressed using human protein homologues where expression data is better characterized. With the exception of four proteins for which no expression data was available, we found that all detected host proteins are present in pancreatic juice, small intestine, colon, rectum and stomach, with particularly high numbers expressed in human pancreatic juice, and the colon (table 1). Pancreatic juice neutralizes stomach acid as it passes from the stomach into the small intestine, and contains multiple digestive enzymes [51]. The detection of proteins expressed in the stomach, pancreatic juice, and the small intestine was therefore consistent with a digestive proteome, while proteins expressed in the colon and rectum were expected in a faecal proteome and supported a host origin. Together, these results indicate the potential for GI proteomes to be preserved within palaeofaeces. Future characterization of expression profiles of dog proteins will improve on these findings. Differences in protein expression between dogs and humans have been observed including in saliva [52], while variation may also exist between different dog lineages. For example, some non-arctic dogs display increased copy numbers of the *AMY2B* gene responsible for the production of amylase, the primary enzyme involved in starch digestion [53].

(iii) Keratins

Keratins are ubiquitous contaminants in shotgun proteomic studies, often derived from handling or the presence of woolen fibres in the sampling or laboratory environment [54], which is exemplified in a previous study of palaeofaeces which detected primarily human keratins [32]. As a result, keratins are frequently excluded from further analysis. Nevertheless, our analysis revealed the presence of keratins and other intermediate filament (IF) proteins which could be attributed to *Canis lupus familiaris* in Nun-3, and to Canidae in Nun-1 and Nun-2 (electronic supplementary material, table S3). Peptides assigned to these proteins show evidence of deamidation, while the presence of ‘pacman peptides’ (peptides with fragmentation at multiple non-tryptic sites) may also indicate protein diagenesis consistent with ancient proteins (but see [55]). Keratins are present in intestinal epithelial cells [56] and are therefore likely to end up in faecal samples as ingested material passes through the digestive tract. In addition, hair was macroscopically observed in the palaeofaeces during sampling and may have originated either from the host or another dog, most likely ingested in relation to grooming. Finally, we note that keratins and IF proteins are potentially informative for dietary reconstruction from dog palaeofaeces as dogs are commonly fed on scraps, such as animal hide, not suitable for human consumption. At present, however, we do not have a secure strategy for distinguishing putative contaminant keratins from potential endogenous proteins that may be informative for diet.

(c) Dietary proteins

Shotgun proteomic data revealed the presence of numerous proteins likely to originate from consumed food. Food-derived proteins were dominated by fishes, and the majority of these were conserved to the level of Salmoninae (electronic supplementary material, table S10). The majority of detected fish

proteins derived from muscle tissue, including myosin motor domain-containing proteins, which could be assigned to the level of Salmoninae and Teleostei in Nun-3 and Nun-4, respectively. Additional fish-derived proteins, which were homologous to actin and titin, both involved in muscle contraction [52], were also detected. Type I collagen (alpha-1(I) and/or alpha-2(I) chains)—a major structural component of connective tissues including bone, tendons and skin—was identified in Nun-1, Nun-2 and Nun-4, and in all cases could be assigned to Salmoninae. Although type I collagen can be used to obtain species-level information through ZooMS and/or LC-MS/MS analysis of bone material, the coverage obtained in the palaeofaeces LC-MS/MS dataset was not sufficient to assign these to a level below Salmoninae.

We identified fish vitellogenin, an egg storage protein surprisingly common in ancient samples [53–56], in Nun-2, Nun-3 and Nun-4. We also detected alpha-1-antiproteinase-like protein from Salmoninae in Nun-2 and Nun-4, and alpha-1,4 glucan phosphorylase from Teleostei in Nun-4. These two proteins are digestive enzymes, which probably originated from the consumption of fish guts. At arctic and subarctic sites like Nunalleq, successful occupation depended on the effective storage of resources that were seasonally available. While bone and muscle proteins may have derived from stored fishes, the presence of proteins associated with fish intestines suggests that some of the analysed palaeofaeces were deposited during the seasonal salmon run. During this time, caught salmon would have been gutted prior to being prepared for human consumption or storage. This is consistent with ethnographic accounts from northern Alaska, which have documented that dogs commonly ate intestines and roe in the summer [57].

These findings by LC-MS/MS analysis are complemented by ZooMS analysis applied to small bone fragments found within some of the palaeofaeces. Collagen spectra were obtained from five bone fragments from Nun-2 and two fragments from Nun-4 (electronic supplementary material, tables S6–S9). Two fragments from Nun-4 dissolved during pretreatment, while another was found to be misidentified and was probably a piece of wood. Two methods of pretreatment using acid demineralisation (method A) and ammonium bicarbonate buffer (AmBic, method B) were applied. Method A generated clearer spectra than method B (electronic supplementary material, tables S6–S9) suggesting that demineralization with acid to release a larger fraction of preserved collagen is preferable in cases where the bones have gone through a digestive tract prior to ZooMS and thus may be more poorly preserved. The bone fragments recovered from Nun-2 were identified as belonging to coho (*Oncorhynchus kisutch*), chinook (*Oncorhynchus tshawytscha*) and chum salmon (*Oncorhynchus keta*) (electronic supplementary material, table S6, tables S8 and S9). From Nun-4, one bone fragment was identified as sockeye salmon (*Oncorhynchus nerka*), and the last assigned to Canidae (electronic supplementary material, tables S6 and S7). These results further demonstrate that ZooMS analyses can successfully be applied to semi-digested bone fragments which are frequently preserved in dog and human palaeofaeces [58] and may provide complementary insight into diet [42]. While ZooMS can provide species-level information for Pacific salmonids, higher resolution analyses, such as LC-MS/MS may be required for this level of taxonomic resolution for mammal species [59].

Previously, indirect dietary evidence from zooarchaeological [7,16] and lipid residue analyses [60] as well as direct dietary evidence obtained from stable isotope analyses from humans [36,61,62] and dogs [7], have indicated a mixed economy at Nunalleq relying predominantly on salmonids, but also marine mammals and terrestrial herbivores. Even with a limited number of analysed samples, the ZooMS and LC-MS/MS analyses corroborated the reliance on salmonids as dog food. This consisted of a range of available salmon species, which may have been fed to dogs without distinction based on the multiple species identified in the Nun-2 sample (electronic supplementary material, table S6). Our findings therefore suggest that the reliance on chum salmon, the so-called 'dog salmon', as dog food, as previously suggested by local traditional knowledge [7], is a later adaptation, potentially related to the commercial value of the species. Interestingly, one bone fragment from Nun-4 was identified as canid, indicating that Nunalleq dogs chewed the bones of wolves, foxes or other dogs, supporting previous observations of dog gnaw marks on discarded dog bones [16]. As tissues from other canids (and especially canid bones) may form part of the diet of domestic dogs, observations of proteins assigned as 'host' may also result from dietary sources. In particular, collagen type I and chondroadherin are abundant in bone and cartilaginous tissues, respectively, and thus, identified peptides from these proteins may also derive from dietary sources, in addition to the host.

Ingested material normally passes through the digestive tract in a matter of days, and as a result, it is highly unlikely that every dietary source of the Nunalleq dogs has been detected here, especially considering the highly seasonal environment [16]. In the future, a more comprehensive study involving additional archaeological contexts and sites could provide important insight into the diversity of feeding strategies across arctic contexts, and beyond.

(d) Challenges and future directions

This study demonstrates the use of palaeoproteomics in identifying the host species of palaeofaeces in cases where morphology is ambiguous, and further showcases the successful sequence identification of components of the host GI proteome. We also confirm that palaeoproteomics and ZooMS can provide high-resolution insight into short-term dietary intake. Nevertheless, given the novelty of this approach, there are several challenges and complexities to be considered in this and future studies.

(i) Imprecise taxonomy and database biases

First, the highly conserved nature of some proteins makes it impossible to assign them to precise taxonomic units (electronic supplementary material, table S2), and differentiating between host proteins and elements of diet can be challenging. Thus, proteins assigned to mammalia or a higher taxonomic classification could derive from either the diet or the dog host. For example, mammalian haemoglobin detected in Nun-3 (electronic supplementary material, table S2) may derive from the dog host, especially as haemoglobin assigned to Carnivora was also identified in the sample. However, this protein could also have originated from caribou or seal, which has been reported from zooarchaeological evidence at Nunalleq (electronic supplementary material, table S11) [63]. Likewise, keratins assigned to Pecora (electronic

supplementary material, table S3) may have originated from caribou hide or intestines consumed by the dog that deposited the Nun-3 sample, but can be difficult to authenticate.

A second major bias relates to the database-matching approach, whereby the incompleteness of available databases and the lack of suitable reference proteomes introduce severe constraints on the ability to identify dietary sources as well as the gut microbiome and potential parasites in complex substrates such as palaeofaeces. If peptides cannot be accurately assigned because particular classes of proteins and species are under-represented in the databases, information may be lost or results misleading. As such it is important to make informed and careful decisions on the composition of databases compiled or selected in data analysis strategies, in order to be aware of potential taxonomic biases based on LC-MS/MS data.

The electronic supplementary material, table S11 presents an overview of taxa identified in zooarchaeological assemblages at Nunalleq [63] and the number of protein accessions available for each taxa in Swiss-Prot and UniProtKB (Swiss-Prot and TReMBL combined). It is clear that variation in this representation has potential downstream effects on interpretation; for example we note that walrus, which has been reported as dog food in northern Alaska [17], is represented by greater than 29 500 proteins in UniProt, but only five are represented in Swiss-Prot. Caribou, which has been suggested as a major resource at Nunalleq is not represented in either database. Likewise, seals, another potential dietary source, are also under-represented, like many marine mammals. In an ideal world, a wide taxonomic diversity would be included in the search strategy; however, databases of large sizes can pose significant constraints [64]. These challenges are well documented in modern metaproteomics [65], but in palaeometaproteomics, they are compounded by the need to widen the search space to account for non-tryptic peptides and post-translational modifications resulting from degradation across archaeological timescales.

(ii) Potential and future applications of faecal palaeoproteomics

We also note a number of topics which will have to be addressed by future research. For instance, methodological investigations are needed to understand the survival of proteins in palaeofaeces. Nunalleq is recognized for its exceptional biomolecular preservation owing to permafrost conditions, but faeces is a highly degradable material, and it is uncertain how proteins preserve in samples that are older or from different archaeological contexts [33].

Future studies exploring the palaeoproteomic profiles of the GI tract may have the potential to reveal insights into GI physiology and disease [66]. For example, proteomic profiles in modern stool samples have shown to be distinct in dogs [67] and humans [68] with acute and chronic GI diseases such as inflammatory bowel disease. Faecal proteomics is also used in diagnosing colorectal cancer [69], celiac disease and cystic fibrosis [70], and quantitative methods may be useful for investigating the expression of immune proteins in palaeofaeces as a sign of infection. Such analyses can inform about the health status of humans and animals, and may help researchers track the emergence and prevalence of GI diseases in the past.

Finally palaeoproteomic analyses of palaeofaeces may offer new insight into community-level gut microbiome composition and function [71], which is complementary to

ancient metagenomic approaches [35]. The gut microbiome is otherwise only preserved in mummified remains, and information from palaeofaeces is therefore a valuable contribution to the understanding of microbiome evolution. The gut microbiome is known to impact overall physiological health through a wide range of functions including digestion, defence against pathogens and immune system response [72]. It responds to changes in diet and shows marked differences between wild and captive mammals [73]. Such analyses could be complicated in dogs; however, as they are known to sometimes engage in coprophagy, while ethnographic accounts from northern Alaska indicate that dog diet may consist of caribou and human faeces [17]. While here we have applied a protein extraction which deliberately reduced the bacterial content, future studies may adopt alternative extractions where this microbiological component can be uncovered, in order to reveal insights into the composition and function of ancient gut microbiomes.

4. Conclusion

Through the analysis of dog coprolites from the site of Nunalleq, Alaska, we demonstrate that palaeofaeces is a viable substrate for palaeoproteomic analysis. We recovered a range of caniform proteins associated with pancreatic juice, small intestine, colon, rectum and stomach and show that the preserved proteome profile is reflective of a host digestive proteome. We further demonstrate that palaeofaeces retains direct dietary information that is accessible through metaproteomic analysis, which, complemented by ZooMS on preserved bone fragments within the palaeofaeces, provides novel insights into tissues and species consumed. Our results demonstrate that dogs at Nunalleq, a pre-contact North American Arctic site, consumed a broad selection of salmonid species available in the local environment, including coho, chinook, sockeye and chum, as well as bones from other canids. Our study also highlights several current challenges in the reconstruction of diet through palaeoproteomic analysis of paleofeces and coprolites, including the need for a better understanding of protein preservation in these substrates, as well as the development of more comprehensive reference databases to maximize protein identifications. Optimized methods to characterize host, dietary and microbial protein sources within coprolites and paleofeces

have the potential to provide novel insight into human and non-human animal diet, GI physiology and disease and microbiome composition and function.

Data accessibility. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium (dataset identifier PXD025714) via the MassIVE partner repository (<http://massive.ucsd.edu>; MSV000087339). ZooMS data have been deposited in Zenodo: <https://doi.org/10.5281/zenodo.4290988>.

Authors' contributions. AK.W.R.: conceptualization, data curation, formal analysis, investigation, methodology and writing—original draft; J.H.: formal analysis, investigation, supervision, writing—original draft, writing—review and editing; K.K.R.: formal analysis, methodology, writing—review and editing; E.M.-M.: conceptualization, resources, writing—review and editing; K.B.: conceptualization, resources, writing—review and editing; M.M.: formal analysis, methodology, writing—review and editing; K.M.: formal analysis, writing—review and editing; M.C.: conceptualization, validation, writing—review and editing; E.C.: conceptualization, supervision, writing—review and editing; C.S.: conceptualization, methodology, supervision, writing—review and editing

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

Competing interests. The authors declare no competing interests.

Funding. This project has received funding from the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under Grant Agreement no. 676154 and the Danish National Research Foundation award PROTEIOS (DNRF128). The onsite collection of samples was carried out by staff and students from the University of Aberdeen, volunteer excavators and the residents of Quinhagak. Fieldwork/primary sampling was funded by an AHRC (AH/K006029/1) grant awarded to Rick Knecht (Aberdeen), K.B. and Charlotta Hillerdal (Aberdeen) and an AHRC-LabEx award (AH/N504543/1) to K.B., Rick Knecht (Aberdeen), Keith Dobney (Liverpool) and Isabelle Sidéra (Nanterre). We had logistical and planning support for fieldwork by Qanirtuuq Incorporated, Quinhagak, Alaska, and the people of Quinhagak, who we also thank for sampling permissions. We thank Prof. Jesper Velgaard Olsen at the Novo Nordisk Center for Protein Research for providing access and resources, which was also funded in part by a donation from the Novo Nordisk Foundation (grant no. NNF14CC0001). Additional support was provided through the New Frontiers in Research grant no. (SSHRC NFRFE-2018-00066).

Acknowledgements. We further acknowledge Takumi Tsutaya for valuable insight in implementing the protein extraction protocol and subsequent discussions, and Samantha Preslee for completing the ZooMS extractions and MALDI-TOF analysis in conjunction with the Technology Facility at the University of York. We are grateful to our three anonymous referees for their constructive comments and suggestions on our original manuscript.

References

1. Frantz LAF *et al.* 2016 Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228–1231. (doi:10.1126/science.aaf3161)
2. Thalmann O *et al.* 2013 Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**, 871–874. (doi:10.1126/science.1243650)
3. Benecke N. 1987 Studies on early dog remains from northern Europe. *J. Archaeol. Sci.* **14**, 31–49. (doi:10.1016/S0305-4403(87)80004-3)
4. Losey RJ, Nomokonova T, Fleming L, Latham K, Harrington L. 2018 Domestication and the embodied human-dog relationship: archaeological perspectives from Siberia. In *Dogs in the north: stories of co-operation and co-domestication* (eds RJ Losey, RP Wishart, JP Laurens Looovers), pp. 8–27. Oxford, UK: Routledge.
5. Bethke B, Burt A. 2020 *Dogs: archaeology beyond domestication*. Gainesville, FL: University Press of Florida.
6. Guiry EJ. 2012 Dogs as analogs in stable isotope-based human paleodietary reconstructions: a review and considerations for future use. *J. Archaeol. Method Theory* **19**, 351–376. (doi:10.1007/s10816-011-9118-z)
7. McManus-Fry E, Knecht R, Dobney K, Richards MP, Britton K. 2018 Dog-human dietary relationships in Yup'ik western Alaska: the stable isotope and zooarchaeological evidence from pre-contact Nunalleq. *J. Archaeol. Sci. Rep.* **17**, 964–972. (doi:10.1016/j.jasrep.2016.04.007)
8. Perri AR, Koster JM, Otárola-Castillo E, Burns JL, Cooper CG. 2019 Dietary variation among indigenous Nicaraguan horticulturalists and their dogs: an ethnoarchaeological application of the Canine Surrogacy Approach. *J. Anthropol. Archaeol.* **55**, 101066. (doi:10.1016/j.jaa.2019.05.002)

9. Arendt M, Cairns KM, Ballard JWO, Savolainen P, Axelsson E. 2016 Diet adaptation in dog reflects spread of prehistoric agriculture. *Heredity* **117**, 301–306. (doi:10.1038/hdy.2016.48)
10. Ní Leathlobhair M *et al.* 2018 The evolutionary history of dogs in the Americas. *Science* **361**, 81–85. (doi:10.1126/science.aao4776)
11. Perri AR, Feuerborn TR, Frantz LAF, Larson G, Malhi RS, Meltzer DJ, Witt KE. 2021 Dog domestication and the dual dispersal of people and dogs into the Americas. *Proc. Natl Acad. Sci. USA* **118**, e2010083118. (doi:10.1073/pnas.2010083118)
12. Morey DF, Aaris-Sørensen K. 2002 Paleoeskimo dogs of the eastern Arctic. *Arctic* **55**, 44–56. (doi:10.14430/arctic689)
13. Howse L. 2019 Hunting technologies and archaeofaunas: societal differences between hunter-gatherers of the eastern Arctic. *J. Archaeol. Method Theory* **26**, 88–111. (doi:10.1007/s10816-018-9366-2)
14. Ameen C *et al.* 2019 Specialized sledge dogs accompanied Inuit dispersal across the North American Arctic. *Proc. R. Soc. B* **286**, 20191929. (doi:10.1098/rspb.2019.1929)
15. Lantis M. 1980 Changes in the Alaskan Eskimo relation of man to dog and their effect on two human diseases. *Arctic Anthropol.* **17**, 1–25.
16. Masson-MacLean E, McManus-Fry E, Britton K. 2020 The archaeology of dogs at the precontact Yup'ik site of Nunalleq, western Alaska. In *Dogs: archaeology beyond domestication* (eds B Bethke, A Burt), pp. 72–102. Gainesville, FL: University Press of Florida.
17. Spencer RF. 1959 *The north Alaskan Eskimo: a study in ecology and society*. Washington: WA: U.S. Government Printing Office.
18. Byers DA, Yesner DR, Broughton JM, Coltrain JB. 2011 Stable isotope chemistry, population histories and Late Prehistoric subsistence change in the Aleutian Islands. *J. Archaeol. Sci.* **38**, 183–196. (doi:10.1016/j.jas.2010.09.005)
19. West CF, France CA. 2015 Human and Canid dietary relationships: comparative stable isotope analysis from the Kodiak Archipelago, Alaska. *J. Ethnobiol.* **35**, 519–535. (doi:10.2993/etbi-35-03-519-535.1)
20. Guiry EJ, Grimes V. 2013 Domestic dog (*Canis familiaris*) diets among coastal Late Archaic groups of northeastern North America: a case study for the canine surrogacy approach. *J. Anthropol. Archaeol.* **32**, 732–745. (doi:10.1016/j.jaa.2013.04.005)
21. Harris AJT, Elliott DA, Guiry EJ, Von Tersch M, Rankin L, Whitridge P, Alexander M, Eriksson G, Grimes V. 2020 Diversity in Labrador Inuit sled dog diets: insights from $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ analysis of dog bone and dentine collagen. *J. Archaeol. Sci. Rep.* **32**, 102424. (doi:10.1016/j.jasrep.2020.102424)
22. Harris AJT *et al.* 2020 Archives of human-dog relationships: genetic and stable isotope analysis of Arctic fur clothing. *J. Anthropol. Archaeol.* **59**, 101200. (doi:10.1016/j.jaa.2020.101200)
23. Hendy J *et al.* 2018 Proteomic evidence of dietary sources in ancient dental calculus. *Proc. R. Soc. B* **285**, 20180977. (doi:10.1098/rspb.2018.0977)
24. Charlton S, Ramsøe A, Collins M, Craig OE, Fischer R, Alexander M, Speller CF. 2019 New insights into Neolithic milk consumption through proteomic analysis of dental calculus. *Archaeol. Anthropol. Sci.* **11**, 6183–6196. (doi:10.1007/s12520-019-00911-7)
25. Wilkin S *et al.* 2020 Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years. *Nat. Ecol. Evol.* **4**, 346–355. (doi:10.1038/s41559-020-1120-y)
26. Maixner F *et al.* 2018 The Iceman's last meal consisted of fat, wild meat, and cereals. *Curr. Biol.* **28**, 2348–2355. (doi:10.1016/j.cub.2018.05.067)
27. Tsutaya T, Mackie M, Koenig C, Sato T, Weber AW, Kato H, Olsen JV, Cappellini E. 2019 Palaeoproteomic identification of breast milk protein residues from the archaeological skeletal remains of a neonatal dog. *Sci. Rep.* **9**, 12841. (doi:10.1038/s41598-019-49183-0)
28. Buckley M, Melton ND, Montgomery J. 2013 Proteomics analysis of ancient food vessel stitching reveals >4000-year-old milk protein. *Rapid Commun. Mass Spectrom.* **27**, 531–538. (doi:10.1002/rcm.6481)
29. Xie M, Shevchenko A, Wang B, Shevchenko A, Wang C, Yang Y. 2016 Identification of a dairy product in the grass woven basket from Gumugou Cemetery (3800 BP, northwestern China). *Quat. Int.* **426**, 158–165. (doi:10.1016/j.quaint.2016.04.015)
30. Yang Y, Shevchenko A, Knaust A, Abuduresule I, Li W, Hu X, Wang C, Shevchenko A. 2014 Proteomics evidence for kefir dairy in Early Bronze Age China. *J. Archaeol. Sci.* **45**, 178–186. (doi:10.1016/j.jas.2014.02.005)
31. Santiago-Rodríguez TM, Narganes-Storde YM, Chanlatte L, Crespo-Torres E, Toranzos GA, Jimenez-Flores R, Hamrick A, Cano RJ. 2013 Microbial communities in pre-columbian coprolites. *PLoS ONE* **8**, e65191. (doi:10.1371/journal.pone.0065191)
32. Shillito L-M, Blong JC, Green EJ, van Asperen EN. 2020 The what, how and why of archaeological coprolite analysis. *Earth-Sci. Rev.* **207**, 103196. (doi:10.1016/j.earscirev.2020.103196)
33. Romaniuk AA *et al.* 2020 Combined visual and biochemical analyses confirm depositor and diet for Neolithic coprolites from Skara Brae. *Archaeol. Anthropol. Sci.* **12**, 274. (doi:10.1007/s12520-020-01225-9)
34. Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Pääbo S. 2000 A molecular analysis of ground sloth diet through the last glaciation. *Mol. Ecol.* **9**, 1975–1984. (doi:10.1046/j.1365-294X.2000.01106.x)
35. Rifkin RF *et al.* 2020 Multi-proxy analyses of a mid-15th century Middle Iron Age Bantu-speaker palaeo-faecal specimen elucidates the configuration of the 'ancestral' sub-Saharan African intestinal microbiome. *Microbiome* **8**, 62. (doi:10.1186/s40168-020-00832-x)
36. Britton K, Knecht R, Nehlich O, Hillerdal C, Davis RS, Richards MP. 2013 Maritime adaptations and dietary variation in prehistoric western Alaska: stable isotope analysis of permafrost-preserved human hair. *Am. J. Phys. Anthropol.* **151**, 448–461. (doi:10.1002/ajpa.22284)
37. Tsutaya T, Mackie M, Sawafuji R, Miyabe-Nishiwaki T, Olsen JV, Cappellini E. In press. Faecal proteomics as a novel method to study mammalian behaviour and physiology. *Mol. Ecol. Resour.* (doi:10.1111/1755-0998.13380)
38. Gurdeep SR, Tanca A, Palomba A, Van der Jeugt F, Verschaffel P, Uzzau S, Martens L, Dawyndt P, Mesuere B. 2019 Unipept 4.0: functional analysis of metaproteome data. *J. Proteome Res.* **18**, 606–615. (doi:10.1021/acs.jproteome.8b00716)
39. Szklarczyk D *et al.* 2019 STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613. (doi:10.1093/nar/gky1131)
40. Stelzer G *et al.* 2016 The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* **54**, 1.30.1–1.30.33. (doi:10.1002/cpbi.5)
41. Buckley M, Collins M, Thomas-Oates J, Wilson JC. 2009 Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 3843–3854. (doi:10.1002/rcm.4316)
42. Welker F, Soressi M, Rendu W, Hublin J-J, Collins M. 2015 Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *J. Archaeol. Sci.* **54**, 279–286. (doi:10.1016/j.jas.2014.12.010)
43. Gibb S, Strimmer K. 2012 MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* **28**, 2270–2271. (doi:10.1093/bioinformatics/bts447)
44. Strohalm M, Hassman M, Kosata B, Kodicek M. 2008 mMass data miner: an open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* **22**, 905–908. (doi:10.1002/rcm.3444)
45. Buckley M, Fraser S, Herman J, Melton ND, Mulville J, Pálsdóttir AH. 2014 Species identification of archaeological marine mammals using collagen fingerprinting. *J. Archaeol. Sci.* **41**, 631–641. (doi:10.1016/j.jas.2013.08.021)
46. Buckley M, Whitcher Kansa S, Howard S, Campbell S, Thomas-Oates J, Collins M. 2010 Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J. Archaeol. Sci.* **37**, 13–20. (doi:10.1016/j.jas.2009.08.020)
47. Kirby DP, Buckley M, Promise E, Trauger SA, Holdcraft TR. 2013 Identification of collagen-based materials in cultural heritage. *Analyst* **138**, 4849–4858. (doi:10.1039/c3an00925d)
48. Welker F *et al.* 2016 Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl Acad. Sci. USA* **113**, 11 162–11 167. (doi:10.1073/pnas.1605834113)
49. Richter KK *et al.* 2020 What's the catch? Archaeological application of rapid collagen-

- based species identification for Pacific Salmon. *J. Archaeol. Sci.* **116**, 105116. (doi:10.1016/j.jas.2020.105116)
50. Somaratne G, Ferrua MJ, Ye A, Nau F, Flourey J, Dupont D, Singh J. 2020 Food material properties as determining factors in nutrient release during human gastric digestion: a review. *Crit. Rev. Food Sci. Nutr.* **60**, 3753–3769. (doi:10.1080/10408398.2019.1707770)
51. Bhagavan NV, Ha C-E. 2015 Chapter 11 - Gastrointestinal digestion and absorption. In *Essentials of medical biochemistry*, 2nd edn (eds NV Bhagavan, C-E Ha), pp. 137–164. San Diego, CA: Academic Press.
52. Leonard TR, Herzog W. 2010 Regulation of muscle force in the absence of actin-myosin-based cross-bridge interaction. *Am. J. Physiol. Cell Physiol.* **299**, C14–C20. (doi:10.1152/ajpcell.00049.2010)
53. Kuckova S, Hynek R, Kodicek M. 2007 Identification of proteinaceous binders used in artworks by MALDI-TOF mass spectrometry. *Anal. Bioanal. Chem.* **388**, 201–206. (doi:10.1007/s00216-007-1206-2)
54. Fremout W, Kuckova S, Chrova M, Sanyova J, Saverwyns S, Hynek R, Kodicek M, Vandenabeele P, Moens L. 2011 Classification of protein binders in artist's paints by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry: an evaluation of principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA). *Rapid Commun. Mass Spectrom.* **25**, 1631–1640. (doi:10.1002/rcm.5027)
55. van der Werf ID, Calvano CD, Laviano R, Simonetti A, Sabbatini L. 2013/1 Multi-technique chemical characterisation of a 12–13th-century painted Crucifix. *Microchem. J.* **106**, 87–94. (doi:10.1016/j.microc.2012.05.011)
56. Shevchenko A, Schuhmann A, Thomas H, Wetzel G. 2018 Fine Endmesolithic fish caviar meal discovered by proteomics in foodcrusts from archaeological site Friesack 4 (Brandenburg, Germany). *PLoS ONE* **13**, e0206483. (doi:10.1371/journal.pone.0206483)
57. Jenness A. 1970 *Dwellers of the Tundra: life in an Alaskan Eskimo village*. New York: NY: Crowell-Collier Press.
58. Reinhard KJ, Bryant VM. 1992 Coprolite analysis: a biological perspective on archaeology. *Archaeol. Method Theory* **4**, 245–288.
59. Rütther PL *et al.* 2021 SPIN - Species by Proteome Investigation. *bioRxiv*. 2021.02.23.432520. (doi:10.1101/2021.02.23.432520)
60. Farrell TFG, Jordan P, Taché K, Lucquin A, Gibbs K, Jorge A, Britton K, Craig OE, Knecht R. 2014 Specialized processing of aquatic resources in prehistoric Alaskan Pottery? A lipid-residue analysis of ceramic sherds from the Thule-Period site of Nunalleq, Alaska. *Arctic Anthropol.* **51**, 86–100. (doi:10.3368/aa.51.1.86)
61. Britton K, McManus-Fry E, Nehlich O, Richards M, Ledger PM, Knecht R. 2018 Stable carbon, nitrogen and sulphur isotope analysis of permafrost preserved human hair from rescue excavations (2009, 2010) at the precontact site of Nunalleq, Alaska. *J. Archaeol. Sci.* **17**, 950–963. (doi:10.1016/j.jasrep.2016.04.015)
62. Britton K. 2019 New isotope evidence for diachronic and site-spatial variation in precontact diet during the Little Ice Age at Nunalleq, southwest Alaska. *Etud. Inuit* **43**, 223. (doi:10.7202/1072717ar)
63. Masson-MacLean E, Houmard C, Knecht R, Sidéra I, Dobney K, Britton K. 2020 Pre-contact adaptations to the Little Ice Age in southwest Alaska: new evidence from the Nunalleq site. *Quat. Int.* **549**, 130–141. (doi:10.1016/j.quaint.2019.05.003)
64. Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, Griffin TJ. 2013 A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. *Proteomics* **13**, 1352–1357. (doi:10.1002/pmic.201200352)
65. Muth T, Benndorf D, Reichl U, Rapp E, Martens L. 2013 Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. *Mol. Biosyst.* **9**, 578–585. (doi:10.1039/C2MB25415H)
66. Jin P, Wang K, Huang C, Nice EC. 2017 Mining the fecal proteome: from biomarkers to personalised medicine. *Expert Rev. Proteomics* **14**, 445–459. (doi:10.1080/14789450.2017.1314786)
67. Honneffer JB, Minamoto Y, Suchodolski JS. 2014 Microbiota alterations in acute and chronic gastrointestinal inflammation of cats and dogs. *World J. Gastroenterol.* **20**, 16 489–16 497. (doi:10.3748/wjg.v20.i44.16489)
68. Vaiopoulou A, Gazouli M, Theodoropoulos G, Zografos G. 2012 Current advantages in the application of proteomics in inflammatory bowel disease. *Dig. Dis. Sci.* **57**, 2755–2764. (doi:10.1007/s10620-012-2291-4)
69. Martínez-Aguilar J, Chik J, Nicholson J, Semaan C, McKay MJ, Molloy MP. 2013 Quantitative mass spectrometry for colorectal cancer proteomics. *Proteomics Clin. Appl.* **7**, 42–54. (doi:10.1002/prca.201200080)
70. Debyser G *et al.* 2016 Faecal proteomics: a tool to investigate dysbiosis and inflammation in patients with cystic fibrosis. *J. Cyst. Fibros.* **15**, 242–250. (doi:10.1016/j.jcf.2015.08.003)
71. Xiong W, Abraham PE, Li Z, Pan C, Hettich RL. 2015 Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. *Proteomics* **15**, 3424–3438. (doi:10.1002/pmic.201400571)
72. Singh RK *et al.* 2017 Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15**, 73. (doi:10.1186/s12967-017-1175-y)
73. McKenzie VJ *et al.* 2017 The effects of captivity on the mammalian gut microbiome. *Integr. Comp. Biol.* **57**, 690–704. (doi:10.1093/icb/ixc090)