Extended methodology for “Investigating the coprolites of Moynagh Lough using biomolecular approaches”

1. Subsampling

Each coprolite was subsampled in the ancient DNA clean laboratory at the University of York to minimise the chances of contamination. The coprolites were sampled using a modified version of the protocol described by Wood and Wilmshurst (2016). Each coprolite was halved using a rotary blade as mineralised coprolites are very hard. Bone inclusions were dissected from one-half of the coprolite which was then hollowed using a carbide Dremel attachment to generate a powder of the coprolite matrix. The powder generated from each coprolite was homogenised and split into subsamples which were used for the metaproteomic, amino acid, and ancient DNA analyses. The remaining half of the coprolite was preserved for potential future analysis.

2. Ancient DNA Analysis

All 11 coprolites and two soil samples were investigated using ancient DNA analyses to access preserved genetic material to identify the depositors of the coprolites and ascertain information about their lifestyles.

2.1. DNA Extraction

All the procedures described below were carried out in the ancient DNA clean laboratory at the University of York. The DNA extraction method described below was chosen due to the high recovery of short DNA fragments. Approximately 100 mg of powdered concretion was suspended in a lysis buffer comprised of 900 µl of 0.5 M EDTA, 50 µl of 10 mg/mL Proteinase K, and 50 µl of 30% N-Lauroylsarcosine and incubated on a rotator for 18 hours at 37°C. The following day DNA within solution was purified by centrifuging each sample to create a pellet, the supernatant was removed and mixed with a binding buffer (15 mL 3M NaOAc, 2.5 mL NaCl, and 0.5 mL Tween 20 to 500 mL Qiagen PB — modified from Allentoft (2015)), which was centrifuged through a High Pure Extender Assembly column provided as part of the High Pure Viral Nucleic Acid Large Volume Kit (Roche). The bound DNA was washed with PE (Qiagen) and eluted in 60 µl of EBT. As coprolites are likely to contain PCR inhibitors, such as humic acids, the DNA elution was then cleaned up further using the OneStep PCR Inhibitor Removal Kit (Zymo). The extracted DNA was quantified using a High Sensitivity Qubit fluorometer.

2.2. Libraries, Amplification and Sequencing

The DNA extracts were diluted to a concentration of 1 ng/µl and 20 ng of DNA was used to build single-stranded libraries using a modified version of the Santa Cruz method (Kapp et al., 2021). Adapters and single-strand binding proteins were added (at concentration 3) and the libraries were cleaned up using Qiagen Min Elutes. Quantitative PCR was performed using NEBNext® Q5U® Master Mix which contains Q5 High Fidelity DNA Polymerase optimised for amplification of degraded samples. The samples were uniquely dual-indexed using the Q5U enzyme during the indexing PCR and the amplified libraries were purified using 1.2X SPRI beads. The libraries were visualised and quantified using the Agilent 4150 TapeStation and samples were pooled in equimolar proportions. To remove as many short PCR dimers as possible, the pools were subsequently cleaned with 1.1X SPRI beads and sequenced on an Illumina HiSeqX using 150 bp paired-end kit.

2.3. Determining the Depositing Species

A key question in coprolite research is identifying the host or depositor species, and four different approaches were tested in this study. First, CoproID (version 1.1.1.) was implemented to distinguish between *Homo sapiens* and *Canis lupus familiaris* depositors (Borry et al., 2020). CoproID compares the number of reads that map to these species and compares the microbial communities of each coprolite with *H. sapiens* and *C. lupus familiaris* gut metagenomics (Borry et al., 2020).

FastQ Screen (Wingett and Andrews, 2018) was utilised, following the method applied by Teasdale et al. (2015), to determine the genome with the highest number of uniquely mapping hits. Using FastQ Screen four species were compared: two species which were likely depositors — *H. sapiens* and *C. lupus familiaris* — and two species which were important dietary resources at Moynagh Lough and Coppergate (McCormick, 1985; O'Connor, 1989) — *Sus scrofa*, and *Bos taurus*. These dietary species were included in the analysis to investigate how dietary DNA sources would affect the depositor identification using DNA. Unique hits to the nuclear genomes were assessed (*H. sapiens* - NC_012920.1, hs37d5; *C. lupus familiaris* - NC_002008.1, CanFam3.1; *S. scrofa* - NC_000845.1, Sscrofa11.1; *B. taurus* - NC_006853.1, bosTau9). Default mapping settings in FastQ Screen were used.

3. Palaeoproteomics

In this study, three palaeoproteomic approaches were applied. First, ZooMS was adopted to gather dietary information. Second, metaproteomics was conducted to attempt to access the host proteome. Thirdly, the amino acid composition of coprolites was characterised for the first time to investigate protein degradation.

3.1. Identifying Bone Inclusions Using Zooarchaeology and ZooMS

Bone inclusions were isolated during the dissection of the coprolites. Despite their small size, the bone fragments were assessed morphologically and compared to modern reference collections at the University of York by Dr Mik Lisowski.

Molecular species identification of the bone inclusions was attempted through collagen peptide mass fingerprinting or Zooarchaeology by Mass Spectrometry (ZooMS) using an acid demineralisation step (Buckley et al., 2009). Once demineralised, the proteins released into solution were cleaved into peptides at lysine and arginine residues using a trypsin incubation at 37°C for 18 hours. This reaction was terminated using 5% TFA before the peptides were cleaned-up using C18 ZipTips (ThermoFisher Scientific). Aliquots were spotted on a stainless steel Matrix-Assisted Laser Desorption Ionization (MALDI) target plate and the peptides were characterised in triplicate via a MALDI-TOF mass spectrometer.

The resulting spectra were prepared using MALDIquant (Gibb and Strimmer, 2012) and visualised and analysed in mMass (Strohalm et al., 2008). Using mMass, the three spectra produced from each sample were averaged, and the resulting averaged spectra were cropped to between 800 and 3500 *m/z*. Peaks were compared to a list of published reference markers, compiled by Dr Samantha Presslee which is publicly available via the BioArCh (University of York) proteomics webpage {https://docs.google.com/spreadsheets/d/1ipm9fFFyha8IEzRO2F5zVXIk0ldwYiWgX5pGqETzBco/edit#gid=1005946405}. Markers are distinguished using the nomenclature nicknames proposed by Brown et al. (2021). Specifically, markers generated by: Buckley et al. (2009), Buckley et al. (2010), Buckley and Collins (2011), Kirby et al. (2013), Welker et al. (2016), Buckley et al. (2017), Eda et al. (2020), and Codlin et al. (2022) were used to make species identifications. This work was carried out by myself (laboratory work and data analysis), Dr Samantha Presslee (laboratory work and data analysis), Lydia Beckett (laboratory work), George Fogarty (laboratory work) and Darcie Eastham (laboratory work).

3.2. Metaproteomics

In order to comprehensively explore the survival of proteins within a coprolite matrix, the samples from Moynagh Lough were analysed using three different protein extraction methods (Chourey et al., 2010; based on Tsutaya et al., 2021; van Doorn et al., 2011 respectively). The Tsutaya et al. (2021) protocol was selected as this was the extraction method successfully employed in permafrost contexts by Runge et al. (2021). The Chourey et al. (2010) protocol was designed for soil extractions, this method was tested as soil, like faeces, is expected to have a multi-source proteome. Finally, the modified van Doorn et al. (2011) protocol was tested due to the minimal number of opportunities for protein loss during cleaning steps. All palaeoproteomic laboratory work was carried out in a dedicated clean room at the University of York.

3.2.1. Protein Extraction Method 1

In Protein Extraction Method 1 (PEM1), proteins were extracted using a protocol designed to preferentially enhance the recovery of the non-microbial proteins in a sample via differential centrifugation, whereby the heavier bacterial cells and cell debris would be pelleted and the resulting supernatant carried forward for analysis (Tsutaya et al., 2021). In a modification to the protocol, the microbial pellet formed during differential centrifugation was also processed alongside the non-microbial fraction for protein recovery.

Briefly, ~150 mg of powdered concretion matrix, alongside positive (porcine gelatin) and negative controls, were demineralised in 10% (v/v) hydrochloric acid for 18 hours at 4°C, a protein pellet was formed by centrifugation and the excess acid was discarded, the pellet was resuspended in Guanidine-HCl (GuHCl) extraction buffer, and incubated at 99°C for 10 minutes. The supernatant was collected and centrifuged for 90 minutes to pellet the microbial content. The pellet (theoretically containing the microbial portion) was rehydrated and processed in the same manner. The proteins in supernatant non-bacterial fraction (*a*) and the bacterial fraction (*b*) were cleaned using C8 SPE cartridges (ThermoFisher Scientific), and the eluate was completely dried. The pellets were resuspended and incubated before a BCA protein quantification assay (Pierce, ThermoFisher Scientific) was completed. Samples in fraction *a* were diluted 1:2 to reduce the protein concentration. Proteins were cleaved using a 3-hour incubation with rLysC at 37°C, followed by an 18-hour trypsin incubation, also at 37°C. Samples were acidified using 5% trifluoroacetic acid (TFA) and desalted using C18 StageTips (Fisher Scientific). The eluates retained a strong yellow colouration. To ensure this colour was not due to particulates in solution, each sample was centrifuged for 10 minutes at 13,000 (rpm) in a benchtop centrifuge and then 50 µL was removed from the upper portion of the supernatant and characterised via Liquid

Chromatography with tandem mass spectrometry (LC-MS/MS) on a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Section 5.2.4.2.4.).

3.2.2. Protein Extraction Method 2

Protein Extraction Method 2 (PEM2) is based on the “Direct Soil Protein Extraction Method (SDS-TCA)” published by Chourey et al. (2010) and was designed to maximise the recovery of unique proteins from soil samples following LC-MS/MS characterization. Whilst soil and faeces are fundamentally different, they both exhibit a complex metaproteomic signature with many potential sources. The method is based on a thermally assisted detergent-based cellular lysis, followed by a TCA precipitation, an SP3 buffer exchange, rLysC and trypsin digestion before desalting and running on an LC-MS/MS. Five samples, taken to be representative of the 12 samples in this study were selected (Figure 5.3 specifies which samples PEM2 was applied to) alongside a positive (porcine gelatin) and negative control were processed in PEM2. Briefly, 50 mg of each powered sample was suspended in an alkaline-SDS buffer, this was heated for 10 minutes at 99°C, cooled to room temperature and centrifuged. The clear supernatant was transferred to new tubes, the proteins were precipitated using chilled TCA, and the samples were rotated at 4°C overnight. Samples were centrifuged for 20 minutes and the supernatant was discarded. The pellet was washed with chilled acetone and resuspended in a guanidine buffer. The SP3 protein isolation method (Hughes et al., 2019) was used to exchange the proteins into ammonia bicarbonate. Proteins were cleaved using 0.4 µg of rLysC and 0.4 µg of trypsin during an 18-hour incubation at 37°C. This reaction was terminated and the peptides were acidified using 5% TFA and the proteins were purified using C18 ZipTips before being characterised by LC-MS/MS (Section 5.2.4.2.4.).

3.2.3. Protein Extraction Method 3

Protein Extraction Method 3 (PEM3) is based on the minimally-invasive ZooMS method presented by van Doorn et al. (2011) but here PEM3 was enhanced to target all proteins (rather than exclusively targeting collagen) by adding a reduction and alkylation step. PEM3 was undertaken on all 12 Moynagh Lough samples, but LC-MS/MS data was only generated for four samples (Figure 5.3).

In brief, ~100 mg of powdered sample was left to soak in 600 µl of ammonium bicarbonate (50 mM) for approximately 76 hours to release soluble proteins into solution. The samples were centrifuged for 10 minutes and the supernatant was collected. The proteins within the supernatant were reduced and alkylated during a 10-minute incubation using a TCEP/CAA solution. As the gastrointestinal tract is filled with digesting enzymes to process food, the fraction was split at this point, half was incubated with 0.4 µg rLysC and trypsin overnight at 37°C (referred to as PEM3), whereas the other half had no enzymatic treatment (referred to as PEM3nd). The enzymatic digest was stopped with 5% TFA and the peptides in both portions were isolated using C18 ZipTips. An aliquot was spotted onto a MALDI plate and screened for LC-MS/MS potential on a MALDI-TOF mass spectrometer. The digested (PEM3) and non-digested (PEM3nd) fractions of four samples with the most MALDI peaks were selected for LC-MS/MS analysis: ML01, ML03, ML09, ML10 (Appendix 5F).

3.2.4. nLC-MS/MS Analysis

The extracted peptides from all three extraction methods were characterised by technicians at the York Centre of Excellence in Mass Spectrometry. Peptides were re-suspended in aqueous 0.1% trifluoroacetic acid (v/v) then loaded onto an mClass nanoflow UPLC system (Waters) equipped with a nanoEaze M/Z Symmetry 100 Å C_{18}, 5 µm trap column (180 µm x 20 mm, Waters) and a PepMap, 2 µm, 100 Å, C_{18} EasyNano nanocapillary column (75 m x 500 mm, Thermo). The trap wash solvent was aqueous 0.05% (v:v) trifluoroacetic acid and the trapping flow rate was 15 µL/min. The trap was washed for 5 min before switching flow to the capillary column. Separation used gradient elution of two solvents: solvent A, aqueous 0.1% (v:v) formic acid; solvent B, acetonitrile containing 0.1% (v:v) formic acid. The flow rate for the capillary column was 300 nL/min and the column temperature was 40°C. The linear multi-step gradient profile was: 3-10% B over 7 minutes, 10-35% B over 30 minutes, 35-99% B over 5 minutes and then proceeded to wash with 99% solvent B for 4 minutes. The column was returned to initial conditions and re-equilibrated for 15 minutes before subsequent injections.

The nanoLC system was interfaced with an Orbitrap Fusion Tribrid mass spectrometer (Thermo) with an EasyNano ionisation source (Thermo). Positive ESI-MS and MS^2 spectra were acquired using Xcalibur software (version 4.0, Thermo). Instrument source settings were: ion spray voltage, 1,900 V; sweep gas, 0 Arb; ion transfer tube temperature; 275°C. MS1 spectra were acquired in the Orbitrap with: 120,000 resolution, scan range: *m/z* 375-1,500; AGC target, $4e^5$; max fill time, 100 milliseconds. Data dependent acquisition was performed in topN mode using a selection of the 12 most intense precursors with charge states >1. Easy-IC was used for internal calibration. Dynamic exclusion was performed for 50 seconds post precursor selection and a minimum threshold for fragmentation was set at $5e^3$. MS2 spectra were acquired in the Orbitrap with: 30,000 resolution, max fill time, 100 milliseconds, HCD; activation energy: 32 NCE.

3.2.5. Peptide Identifications

Raw files were converted to Mascot generic format and searched against the Swiss-Prot database (database downloaded 14/12/2022) using the University of York's Mascot server (Version 2.8.0) (Perkins et al., 1999) allowing two missed cleavage sites. Carbamidomethyl (C) was a fixed modification, whereas acetylation (N-term), deamidation (NQ), glutamine to pyroglutamic acid (N-term Q), glutamic acid to pyroglutamic acid (N-term E), hydrolysis of lysine (K), oxidation (M), and hydroxylation (P) were included as variable modifications based on the modifications recorded by Runge et al. (2021). The search was semi-tryptic for all extraction attempts, with the exception of PEM3nd which did not include an enzyme digestion step. PEM3nd was searched with no specified enzyme. An FDR rate of 1% was applied and proteins with two or more unique peptide matches were recorded. Peptides matching to trypsin, lysyl endopeptidase and keratin were not recorded. All proteins identified in the blanks were considered exogenous and these peptides were not recorded for the coprolite samples. The taxonomy of peptides of interest was ascertained using Unipept (Mesuere et al., 2012), and unexpected taxonomic identifications were searched using BLASTp (Altschul et al., 1990).

3.3. Detection of Hydrolysed and Free Amino Acids

Like metaproteomics, amino acids were characterised for the 11 coprolites and 1 soil sample from Moynagh Lough. Free amino acids are not usually characterised in open systems (such as coprolites) due to the interaction with the environment (Collins and Riley, 2000), but in order to provide a direct comparison for hydrolysed amino acids, free amino acids were measured here. The sample preparation for amino acid detection was based on the methods outlined in High et al. (2015) and Oonk et al. (2012).

In brief, between 1 - 3 mg of powdered coprolite was: 1) dissolved in 200 µl 2M HCl to test for free amino acids and 2) hydrolysed in 200 µl of 7M HCl, and flushed with nitrogen before incubating at 110°C for 18 hours to quantify amino acids released from peptides by acid hydrolysis. For both experiments, the acid was evaporated entirely using a centrifugal evaporator. The dried residue (containing the amino acids) was rehydrated in a weak acid (0.01 M HCl, 1.5 mM sodium azide) which also contained L-homo-arginine (a non-protein amino acid used as an internal standard). Before being characterised using reverse-phase high-pressure liquid chromatography (RP-HPLC) as described by Kaufman and Manley (1998). The order of retention is as follows: L-Asx, D-Asx, L-Glx, D-Glx, L-Ser, D-Ser, L-Thr, L-His, Gly, L-Arg, D-Arg, L-Ala, L-hArg (internal standard), D-Ala, L-Tyr, D-Tyr, L-Val, L-Met, D-Met, D-Val, L-Phe, L-Ile, D-Phe, L-Leu, D-Ile, D-Leu. The retention times of the amino acids along with the area under each peak were recorded. The peak areas were used to calculate: 1) the amino acid concentration; 2) the racemisation values of the amino acids;

and 3) the overall amino acid percentage composition of each sample. Co-eluting peaks with Tyr, Leu and Ile mean that accurately quantifying concentrations for the D-isomer of these amino acids is difficult (Dickinson et al., 2019), so they were not plotted. Additionally, as it is not possible to obtain racemisation data from Thr, His, Gly and Arg, these amino acids are also not included in that analysis.

The concentration, composition and racemisation values obtained from archaeological bone samples characterised by Presslee et al. (2021) were averaged to create representative values for an archaeological bone. This representative bone profile was used to compare the results from coprolites to another archaeological open system. To investigate the relationship between amino acid racemisation and the preservation of ancient DNA, the obtained racemisation values were plotted alongside the number of mapped depositor reads which were normalised to account for different sequencing depths.

4. Radiocarbon Dating

To investigate the coherence between radiocarbon dates obtained from the matrix of a coprolite and the bone inclusions embedded within it samples were sent to Vilnius Radiocarbon, Estonia, for radiocarbon dating. After subsampling for biomolecular analyses and completing a zooarchaeological and ZooMS assessment, pieces of four coprolites weighing >2 g and bone inclusions from those coprolites were selected for radiocarbon dating. The samples with both coprolites and bone inclusions dated were ML01 and ML05 from Moynagh Lough. Additional dates were obtained from the bone inclusions from 3 samples - specifically, ML02, ML07 and ML10. In all cases, bone fragments dissected from the same coprolite were combined together to increase the collagen yield from the bone. Dates were calibrated and shown as calAD using the OxCal (v4.4.4.) calibration curve (Bronk Ramsey, 1994) which utilises IntCal20 (Reimer et al., 2020).

5. References

Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Della Casa, P., Dąbrowski, P., Duffy, P.R., Ebel, A.V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Price, T.D., Saag, L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V.I., Szeverényi, V., Tóth, G., Trifanova, S.V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Population genomics of Bronze Age Eurasia. Nature 522, 167–172.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Borry, M., Cordova, B., Perri, A., Wibowo, M., Honap, T.P., Ko, J., Yu, J., Britton, K., Girdland-Flink, L., Power, R.C., Stuijts, I., Salazar-García, D.C., Hofman, C., Hagan, R., Kagoné, T.S., Meda, N., Carabin, H., Jacobson, D., Reinhard, K., Lewis, C., Kostic, A., Jeong, C., Herbig, A., Hübner, A., Warinner, C., 2020. CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content. PeerJ 8, e9001.

Bronk Ramsey, C., 1994. Analysis of chronological information and radiocarbon calibration: the program OxCal. Archaeological Computing Newsletter.

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

Buckley, M., Collins, M.J., 2011. Collagen survival and its use for species identification in Holocene-lower Pleistocene bone fragments from British archaeological and paleontological sites. Antiqua.

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

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

Buckley, M., Kansa, S.W., Howard, S., Campbell, S., 2010. Distinguishing between archaeological sheep and goat bones using a single collagen peptide. Journal of.

Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., Brodie, E.L., Hettich, R.L., 2010. Direct cellular lysis/protein extraction protocol for soil metaproteomics. J. Proteome Res. 9, 6615–6622.

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

Collins, M.J., Riley, M., 2000. Amino acid racemization in biominerals, the impact of protein degradation and loss. Oxford University Press.

Dickinson, M.R., Lister, A.M., Penkman, K.E.H., 2019. A new method for enamel amino acid racemization dating: A closed system approach. Quat. Geochronol. 50, 29–46.

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

Gibb, S., Strimmer, K., 2012. MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinformatics 28, 2270–2271.

High, K., Milner, N., Panter, I., Penkman, K.E.H., 2015. Apatite for destruction: investigating bone degradation due to high acidity at Star Carr. J. Archaeol. Sci. 59, 159–168.

Hughes, C.S., Moggridge, S., Müller, T., Sorensen, P.H., Morin, G.B., Krijgsveld, J., 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc. 14, 68–85.

Kapp, J.D., Green, R.E., Shapiro, B., 2021. A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. J. Hered. 112, 241–249.

Kaufman, D.S., Manley, W.F., 1998. A new procedure for determining dl amino acid ratios in fossils using reverse phase liquid chromatography. Quat. Sci. Rev. 17, 987–1000.

Kirby, D.P., Buckley, M., Promise, E., Trauger, S.A., Holdcraft, T.R., 2013. Identification of collagen-based materials in cultural heritage. Analyst 138, 4849–4858.

McCormick, F., 1985. Interim report on the animal bones from Moynagh Lough. Riocht na Midhe.

Mesuere, B., Devreese, B., Debyser, G., Aerts, M., Vandamme, P., Dawyndt, P., 2012. Unipept: Tryptic Peptide-Based Biodiversity Analysis of Metaproteome Samples. J. Proteome Res. 11, 5773–5780.

O'Connor, T.P., 1989. Bones from Anglo-Scandinavian Levels at 16-22 Coppergate. York Archaeological Trust.

Oonk, S., Cappellini, E., Collins, M.J., 2012. Soil proteomics: An assessment of its potential for archaeological site interpretation. Org. Geochem. 50, 57–67.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

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

Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Ramsey, C.B., Butzin, M., Cheng, H., Lawrence Edwards, R., Friedrich, M., Grootes, P.M., Guilderson, T.P., Hajdas, I., Heaton, T.J., Hogg, A.G., Hughen, K.A., Kromer, B., Manning, S.W., Muscheler, R., Palmer, J.G., Pearson, C., van der Plicht, J., Reimer, R.W., Richards, D.A., Marian Scott, E., Southon, J.R., Turney, C.S.M., Wacker, L., Adolphi, F., Büntgen, U., Capano, M., Fahrni, S.M., Fogtmann-Schulz, A., Friedrich, R., Köhler, P., Kudsk, S., Miyake, F., Olsen, J., Reinig, F., Sakamoto, M., Sookdeo, A., Talamo, S., 2020. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). Radiocarbon 62, 725–757.

Runge, A.K.W., Hendy, J., Richter, K.K., Masson-MacLean, E., Britton, K., Mackie, M., McGrath, K., Collins, M., Cappellini, E., Speller, C., 2021. Palaeoproteomic analyses of dog palaeofaeces reveal a preserved dietary and host digestive proteome. Proc. Biol. Sci. 288, 20210020.

Strohalm, M., Hassman, M., Kosata, B., Kodícek, M., 2008. mMass data miner: an open source alternative for mass spectrometric data analysis. Rapid Commun. Mass Spectrom. 22, 905–908.

Teasdale, M.D., van Doorn, N.L., Fiddyment, S., Webb, C.C., O'Connor, T., Hofreiter, M., Collins, M.J., Bradley, D.G., 2015. Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370, 20130379.

Tsutaya, T., Mackie, M., Sawafuji, R., Miyabe-Nishiwaki, T., Olsen, J.V., Cappellini, E., 2021. Faecal proteomics as a novel method to study mammalian behaviour and physiology. Mol. Ecol. Resour. 21, 1808–1819.

van Doorn, N.L., Hollund, H., Collins, M.J., 2011. A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. Archaeol. Anthropol. Sci. 3, 281–289.

Welker, F., Hajdinjak, M., Talamo, S., Jaouen, K., Dannemann, M., David, F., Julien, M., Meyer, M., Kelso, J., Barnes, I., Brace, S., Kamminga, P., Fischer, R., Kessler, B.M., Stewart, J.R., Pääbo, S., Collins, M.J., Hublin, J.-J., 2016. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. Proc. Natl. Acad. Sci. U. S. A. 113, 11162–11167.

Wingett, S.W., Andrews, S., 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res. 7, 1338.

Wood, J.R., Wilmshurst, J.M., 2016. A protocol for subsampling Late Quaternary coprolites for multi-proxy analysis. Quat. Sci. Rev. 138, 1–5.