CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content

Maxime Borry[1], Bryan Cordova[1], Angela Perri[2, 11], Marsha C. Wibowo[17, 18, 3], Tanvi Honap[8, 16], Wing Tung Jada Ko[4], Jie Yu[5], Kate Britton[11, 15], Linus Girdland Flink[15, 19], Robert C. Power[11,12], Ingelise Stuijts[13], Domingo Salazar Garcia[14], Courtney A. Hofman[8,16], Richard W. Hagan[1], Thérèse Samdapawindé Kagone[6], Nicolas Meda[6], Hélène Carabin[7], David Jacobson[8, 16], Karl Reinhard[9], Cecil M. Lewis, Jr.[8, 16], Aleksandar Kostic[17, 18, 3], Choongwon Jeong[1], Alexander Herbig[1], Alexander Hübner[1], and Christina Warinner[1,4,10]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany 07745
[2]Department of Archaeology, Durham University, Durham, UK DH13LE
[3]Harvard Medical School, Department of Microbiology, Boston, MA, USA 02215
[4]Department of Anthropology, Harvard University, Cambridge, MA, USA 02138
[5]Department of History, Wuhan University, Wuhan, China
[6]Centre MURAZ Research Institute/Ministry of Health, Bobo-Dioulasso, Burkina Faso
[7]Département de pathologie et de microbiologie, Faculté de Médecine vétérinaire-Université de Montréal, Saint-Hyacinthe, Canada, QC J2S 2M2
[8]Department of Anthropology, University of Oklahoma, Norman, OK, USA 73019
[9]School of Natural Resources, University of Nebraska, Lincoln, NE, USA 68583
[10]Faculty of Biological Sciences, Friedrich-Schiller University, Jena, Germany, 07743
[11]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[12]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig-Maximilians-Universität München, Munich
[13]The Discovery Programme, 6 Mount Street Lower, Dublin 2, Ireland
[14]Grupo de Investigación en Prehistoria IT-622-13 (UPV- EHU), IKERBASQUE-Basque Foundation for Science
[15]Department of Archaeology, University of Aberdeen, St Mary's Building, Elphinstone Road, Aberdeen, AB24 3UF, UK
[16]Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, Norman, OK, USA 73019
[17]Joslin Diabetes Center, Section on Pathophysiology and Molecular Pharmacology, Boston, MA, USA
[18]Joslin Diabetes Center, Section on Islet Cell and Regenerative Biology, Boston, MA, USA
[19]School of Natural Sciences and Psychology, Liverpool John Moores University, L3 3AF Liverpool, United Kingdom

Corresponding author:

Maxime Borry, Christina Warinner

Email address: borry@shh.mpg.de, warinner@shh.mpg.de

ABSTRACT

Shotgun metagenomics applied to archaeological feces (paleofeces) can bring new insights into the composition and functions of human and animal gut microbiota from the past. However, paleofeces often undergo physical distortions in archaeological sediments, making their source species difficult to identify on the basis of fecal morphology or microscopic features alone. Here we present a reproducible and scalable pipeline using both host and microbial DNA to infer the host source of fecal material. We apply this pipeline to newly sequenced archaeological specimens and show that we are able to distinguish morphologically similar human and canine paleofeces, as well as non-fecal sediments, from a range of archaeological contexts.

INTRODUCTION

The gut microbiome, located in the distal colon and primarily studied through the analysis of feces, is the largest and arguably most influential microbial community within the body (Huttenhower et al., 2012). Recent investigations of the human microbiome have revealed that it plays diverse roles in health and disease, and gut microbiome composition has been linked to a variety of human health states, including inflammatory bowel diseases, diabetes, and obesity (Kho and Lal, 2018). To investigate the gut microbiome, metagenomic sequencing is typically used to reveal both the taxononomic composition (i.e., which bacteria are there) and the functions the microbes are capable of performing (i.e., their potential metabolic activities) (Sharpton, 2014). Given the importance of the gut microbiome in human health, there is great interest in understanding its recent evolutionary and ecological history (Warinner and Lewis Jr, 2015; Davenport et al., 2017).

Paleofeces, either in an organic or partially mineralized (coprolite) state, present a unique opportunity to directly investigate changes in the structure and function of the gut microbiome through time (Warinner et al., 2015). Paleofeces are found in a wide variety of archaeological contexts around the world and are generally associated with localized processes of dessication, freezing, or mineralization. Paleofeces can range in size from whole, intact fecal pieces (Jiménez et al., 2012) to millimeter-sized sediment inclusions identifiable by their high phosphate and fecal sterol content (Sistiaga et al., 2014). Although genetic approaches have long been used to investigate dietary DNA found within human (Gilbert et al., 2008; Poinar et al., 2001) and animal (Poinar et al., 1998; Hofreiter et al., 2000; Bon et al., 2012; Wood et al., 2016) paleofeces, it is only recently that improvements in metagenomic sequencing and bioinformatics have enabled detailed characterization of their microbial communities (Tito et al., 2008, 2012; Warinner et al., 2017).

However, before evolutionary studies of the gut microbiome can be conducted, it is first necessary to confirm the host source of the paleofeces under study. Feces can be difficult to taxonomically assign by morphology alone (Supplementary Note), and human and canine feces can be particularly difficult to distinguish in archaeological contexts (Poinar et al., 2009). Since their initial domestication more than 12,000 years ago (Frantz et al., 2016), dogs have often lived in close association with humans, and it is not uncommon for human and dog feces to co-occur at archaeological sites. Moreover, dogs often consume diets similar to humans because of provisioning or refuse scavenging (Guiry, 2012), making their feces difficult to distinguish based on dietary contents. Even well-preserved fecal material degrades over time, changing in size, shape, and color (Figure 1). The combined analysis of host and microbial ancient DNA (aDNA) within paleofeces presents a potential solution to this problem.

Previously, paleofeces host source has been genetically inferred on the basis of PCR-amplified mitochondrial DNA sequences alone (Hofreiter et al., 2000); however, this is problematic in the case of dogs, which, in addition to being pets and working animals, were also eaten by many ancient cultures (Clutton-Brock and Hammond, 1994; Rosenswig, 2007; Kirch and O'Day, 2003; Podberscek, 2009), and thus trace amounts of dog DNA may be expected to be present in the feces of humans consuming dogs. Additionally, dogs often scavenge on human refuse, including human excrement (Butler and Du Toit, 2002), and thus ancient dog feces could also contain trace amounts of human DNA, which could be further inflated by PCR-based methods.

A metagenomics approach overcomes these issues by allowing a quantitative assessment of eukaryotic DNA at a genome-wide scale, including the identification and removal of modern human contaminant

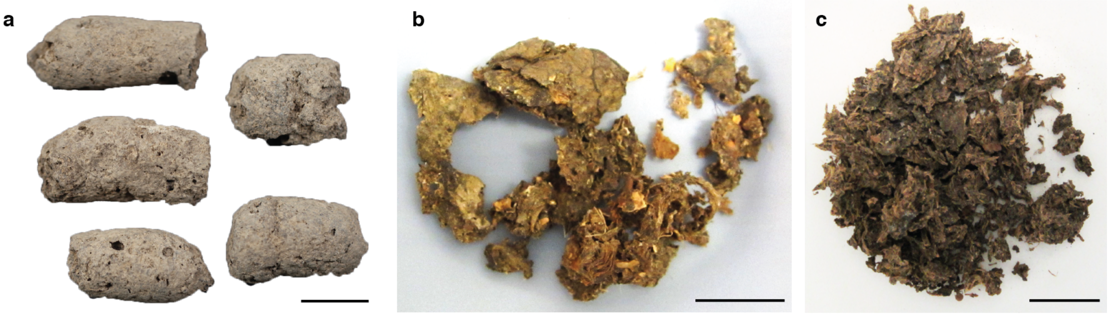


Figure 1. Examples of archaeological paleofeces analyzed in this study.
(**a**) H29-3, from Anhui Province, China, Neolithic period; (**b**) Zape 2, from Durango, Mexico, ca. 1300 BP; (**c**) Zape 28, from Durango, Mexico, ca. 1300 BP. Paleofeces ranged from slightly mineralized intact pieces (a) to more fragmentary organic states (b, c), and color ranged from pale gray (a) to dark brown (c). Each scale bar represents 2 cm.

DNA that could potentially arise during excavation or subsequent curation or storage. It also allows for the microbial composition of the feces to be taken into account. Gut microbiome composition differs among mammal species (Ley et al., 2008), and thus paleofeces microbial composition could be used to confirm and authenticate host assignment. Available microbial tools, such as SourceTracker (Knights et al., 2011) and FEAST (Shenhav et al., 2019), can be used to perform the source prediction of microbiome samples from uncertain sources (sinks) using a reference dataset of source-labeled microbiome samples and, respectively, Gibbs sampling or an Expectation-Maximization algorithm. However, although SourceTracker has been widely used for modern microbiome studies and has even been applied to ancient gut microbiome data (Tito et al., 2012) (Hagan et al., 2019), it was not designed to be a host species identification tool for ancient microbiomes.

In this work we present a bioinformatics method to infer and authenticate the host source of paleofeces from shotgun metagenomic DNA sequencing data: coproID (**copro**lite **ID**entification). coproID combines the analysis of putative host ancient DNA with a machine learning prediction of the feces source based on microbiome taxonomic composition. Ultimately, coproID predicts the host source of a paleofeces specimen from the shotgun metagenomic data derived from it. We apply coproID to previously published modern fecal datasets and show that it can be used to reliably predict their host. We then apply coproID to a set of newly sequenced paleofeces specimens and non-fecal archaeological sediments and show that it can discriminate between feces of human and canine origin, as well as between fecal and non-fecal samples.

MATERIAL AND METHODS

Gut microbiome reference datasets

Previously published modern reference microbiomes were chosen to represent the diversity of potential paleofeces sources and their possible contaminants, namely human fecal microbiomes from Non-Westernized Human/Rural (NWHR), and Westernized Human/Urban (WHU) communities, dog fecal microbiomes, and soil samples (Table 1). Because the human datasets had been filtered to remove human genetic sequences prior to database deposition, we additionally generated new sequencing data from 118 fecal specimens from both NWHR and WHU populations (Table S5) in order to determine the average proportion and variance of host DNA in human feces.

Metagenome source	Food production	N	Analysis	Source
Homo sapiens - USA	WHU	36	microbiome	The Human Microbiome Project Consortium et al. (2012)
Homo sapiens - India (Bhopal and Kerala)	WHU & NWHR	19	microbiome	Dhakan et al. (2019)
Homo sapiens - Fiji (agrarian villages)	NWHR	20	microbiome	Brito et al. (2019)
Homo sapiens - Madagascar	NWHR	110	microbiome	Pasolli et al. (2019)
Homo sapiens - Brazil (Yanomami)	NWHR	3	microbiome	Pasolli et al. (2019)
Homo sapiens - Peru (Tunapuco)	NWHR	12	microbiome	Obregon-Tito et al. (2015)
Homo sapiens - Tanzania (Hadza)	NWHR	38	microbiome	Rampelli et al. (2015)
Homo sapiens - Peru (Matses)	NWHR	24	microbiome	Obregon-Tito et al. (2015)
Homo sapiens - USA (Boston)	WHU	49	host DNA	This study
Homo sapiens - Burkina Faso	NWHR	69	host DNA	This study
Canis familiaris	-	150	microbiome and host DNA	Coelho et al. (2018)
Soil	-	16	microbiome	Fierer et al. (2012)
Soil	-	2	microbiome	CSIR and aromatic plants (2016)
Soil	-	2	microbiome	Orellana et al. (2018)

Table 1. Modern reference microbiome datasets

Archaeological samples

A total of 20 archaeological samples, originating from 10 sites and spanning periods from 7200 BP to the medieval era, were selected for this study. Among these 20 samples, of which 17 are newly sequenced, 13 are paleofeces, 4 are midden sediments, and 3 are sediments obtained from human pelvic bone surfaces. (Table 2).

Archeological ID	Laboratory ID	Site Name	Region	Period	Sample type	Archaeologically suspected species	Plot ID
Zape 2*	ZSM002	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	01
Zape 5*	ZSM005	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	02
Zape 23	ZSM023	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN or CANID	03
Zape 25	ZSM025	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	04
Zape 27	ZSM027	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	05
Zape 28*	ZSM028	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	06
Zape 29	ZSM029	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	07
Zape 31	ZSM031	Cueva de los Muertos Chiquitos	Mexico	1300 BP	Paleofeces	HUMAN	08
H29-1	AHP001	Xiaosungang	China	Neolithic 7200-6800 BP	Paleofeces	CANID or CERVID	09
H35-1	AHP002	Xiaosungang	China	Neolithic 7200-6800 BP	Paleofeces	CANID or CERVID	10
H29-2	AHP003	Xiaosungang	China	Neolithic 7200-6800 BP	Paleofeces	CANID or CERVID	11
H29-3	AHP004	Xiaosungang	China	Neolithic 7200-6800 BP	Paleofeces	CANID or CERVID	12
LG 4560.69	YRK001	Surrey	UK	Post-Medieval	Paleofeces	HUMAN	13
AP3-C197S163	DRL001.A	Derragh	Ireland	Mesolithic	Midden Sediment	-	14
AP4-A6-2860	CBA001.A	Cabeço das Amoreiras	Portugal	Mesolithic	Midden Sediment	-	15
AP5-798-162	BRF001.A	Binchester Roman Fort	England	Roman	Midden Sediment	-	16
AP6-LPZ702	LEI010.A	Leipzig	Germany	10th- 11th century AD	Midden Sediment	-	17
AP7-6-28353	ECO004.D	El Collado	Spain	Mesolithic	Pelvic Sediment	-	18
AP8-CMN-M1	CMN001.D	Cingle del Mas Nou	Spain	Mesolithic	Pelvic Sediment	-	19
AP9-17590	MLP001.A	Molpir	Slovakia	7th century BC	Pelvic Sediment	-	20

*Metagenomic data were previously published in (Hagan et al., 2019)

Table 2. Archaeological samples

Sampling

Paleofeces specimens from Mexico were sampled in a dedicated aDNA cleanroom in the Laboratories for Molecular Anthropology and Microbiome Research (LMAMR) at the University of Oklahoma, USA. Specimens from China were sampled in a dedicated aDNA cleanroom at the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena, Germany. All other specimens were first sampled at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) in Leipzig, Germany before being transferred to the MPI-SHH for further processing. Sampling was performed using a sterile stainless steel spatula or scalpel, followed by homogenization in a mortar and pestle, if necessary. Because the specimens from Xiaosungang, China were very hard and dense, a rotary drill was used to section the coprolite prior to sampling. Where possible, fecal material was sampled from the interior of the specimen rather than the surface. Specimens from Molphir and Leipzig were received suspended in a buffer of trisodium phosphate, glycerol, and formyl following screening for parasite eggs using optical microscopy. For each paleofeces specimen, a total of 50-200 mg was analyzed.

Modern feces were obtained under informed consent from Boston, USA (WHU) (Wibowo et al., 2019) from a long-term (>50 years) type 1 diabetes cohort, and from villages in Burkina Faso (NWHR) as part of broader studies on human gut microbiome biodiversity and health-associated microbial communities. Feces were collected fresh and stored frozen until analysis. A total of 250 mg was analyzed for each fecal specimen,

DNA Extraction

For paleofeces and sediment samples, DNA extractions were performed using a silica spin column protocol (Dabney et al., 2013) with minor modifications in dedicated aDNA cleanrooms located at LMAMR (Mexican paleofeces) and the MPI-SHH (all other paleofeces). At LMAMR, the modifications followed those of protocol D described in (Hagan et al., 2019). DNA extractions at the MPI-SHH were similar, but omitted the initial bead-beating step, and a single silica column was used per sample instead of two. Additionally, to reduce centrifugation errors, DNA extractions performed at the MPI-SHH substituted the column apparatus from the High Pure Viral Nucleic Acid Large Volume Kit (Roche, Switzerland) in place of the custom assembled Zymo-reservoirs coupled to MinElute (Qiagen) columns described in (Dabney et al., 2013). At both locations, non-template negative extraction controls were processed alongside samples to identify and monitor potential contamination.

For modern feces, DNA was extracted from Burkina Faso fecal samples using the AllPrep PowerViral DNA/RNA Qiagen kit at Centre MURAZ Research Institute in Burkina Faso. DNA was extracted from the Boston fecal material using the ZymoBIOMICS DNA Miniprep Kit (D4303) at the Joslin Diabetes Center as described in (Wibowo et al., 2019).

Library preparation and Sequencing

For paleofeces and sediment samples, double-stranded, dual-indexed shotgun Illumina libraries were constructed following (Meyer and Kircher, 2010) using either the NEBNext DNA Library Prep Master Set (E6070) kit (Hagan et al., 2019; Mann et al., 2018) for the Mexican paleofeces or individually purchased reagents (Mann et al., 2018) for all other samples. Following library amplification using a Kapa HiFi Uracil+ polymerase or Agilent Pfu Turbo Cx Hotstart polymerase, the libraries were purified using a Qiagen MinElute PCR Purification kit and quantified using either a BioAnalyzer 2100 with High Sensitivity DNA reagents or an Agilent Tape Station D1000 Screen Tape kit. The Mexican libraries were pooled in equimolar amounts and sequenced on an Illumina HiSeq 2000 using 2x100 bp paired-end sequencing. All other libraries were pooled in equimolar amounts and sequenced on an Illumina HiSeq 4000 using 2x75 bp paired-end sequencing.

For modern NWHR feces, double-stranded, dual-indexed shotgun Illumina libraries were constructed in a dedicated modern DNA facility at LMAMR. Briefly, after DNA quantification using a Qubit dsDNA Broad Range Assay Kit, DNA was sheared using a QSonica Q800R in 1.5mL 4°C cold water at 50% amplitude for 12 minutes to aim for a fragment size between 400 and 600 bp. Fragments shorter than 150 bp were removed using Sera-Mag SpeedBeads and a Alpaqua 96S Super Magnet Plate. End-repair and A-tailing was performed using the Kapa HyperPrep EndRepair and A-Tailing Kit, and Illumina sequencing adapters were added. After library quantification, libraries were dual-indexed in an indexing PCR over four replicates, pooled, and purified using the SpeedBeads. Libraries were quantified using the Agilent Fragment Analyzer, pooled in equimolar ratios, and size-selected using the Pippin Prep to a target size range of 400-600 bp. Libraries were sequenced on an Illumina NovaSeq S1 using 2x150 bp paired-end sequencing at the Oklahoma Medical Research Foundation Next-Generation Sequencing Core facility. Modern WHU libraries were generated using the NEBNext DNA library preparation kit following manufacturer's recommendations, after fragmentation by shearing for a target fragment size of 350 bp as described in (Wibowo et al., 2019). The libraries were then pooled and sequenced by Novogene on a NovaSeq S4 using 2x150 bp paired-end sequencing.

Proportion of host DNA in gut microbiome

Because it is standard practice to remove human DNA sequences from metagenomics DNA sequence files before data deposition into public repositories, we were unable to infer the proportion of human DNA in human feces from publicly available data. To overcome this problem, we measured the proportion of human DNA in two newly generated fecal metagenomics datasets from Burkina Faso (NWHR) and Boston, U.S.A. (WHU) (Table S5). To measure the proportion of human DNA in each fecal dataset, we used the Anonymap pipeline (Borry, 2019a) to perform a mapping with Bowtie 2 (Langmead and Salzberg, 2012) with the parameters `--very-sensitive -N 1` after adapter cleaning and reads trimming for ambiguous and low-quality bases with a QScore below 20 by AdapterRemoval v2 (Schubert et al., 2016). To preserve the anonymity of the donors, the sequences of mapped reads were then replaced by `Ns` thus anonymizing the alignment files. We obtained the proportion of host DNA per sample by dividing the number of mapped reads by the total number of reads in the sample. The proportion of host DNA in dog feces was determined from the published dataset Coelho et al. (2018) as described above, but without the anonymization step.

coproID pipeline

Data were processed using the coproID pipeline v1.0 (Figure 2) (DOI: 10.5281/zenodo.2653757) written using Nextflow (Di Tommaso et al., 2017) and made available through nf-core (Ewels et al., 2019). Nextflow is a Domain Specific Language designed to ensure reproducibility and scalability for scientific pipelines, and nf-core is a community-developed set of guidelines and tools to promote standardization and maximum usability of Nextflow pipelines.

coproID consists of 5 different steps:

Preprocessing

Fastq sequencing files are given as an input. After quality control analysis with FastQC (Andrews et al., 2010), raw sequencing reads are cleaned from sequencing adapters and trimmed from ambiguous and low-quality bases with a QScore below 20, while reads shorter than 30 base pairs are discarded using AdapterRemoval v2. By default, paired-end reads are merged on overlapping base pairs.

Mapping

The preprocessed reads are then aligned to each of the target species genomes (source species) by Bowtie2 with the `--very-sensitive` preset while allowing for a mismatch in the seed search (`-N 1`).

When running coproID with the ancient DNA mode (`--adna`), alignments are filtered by PMDtools (Skoglund et al., 2014) to only retain reads showing post-mortem damages (PMD). PMDtools default settings are used, with specified library type, and only reads with a PMDScore greater than three are kept.

Computing host DNA content

Next, filtered alignments are processed in Python using the Pysam library (pysam developers, 2018). Reads matching above the identity threshold of 0.95 to multiple host genomes are flagged as common reads $reads_{commons}$ whereas reads mapping above the identity threshold to a single host genome are flagged as genome-specific host reads $reads_{spec\ g}$ to each genome g. Each source species host DNA is normalized by genome size and gut microbiome host DNA content such as:

$$NormalizedHostDNA(source\ species) = \frac{\sum length(reads_{spec\ g})}{genome_{g\ length} \cdot endo_{g}} \tag{1}$$

where for each species of genome g, $\sum length(reads_{spec\ g})$ is the total length of all $reads_{spec\ g}$, $genome_{g\ length}$ is the size of the genome, and $endo_g$ is the host DNA proportion in the species gut microbiome.

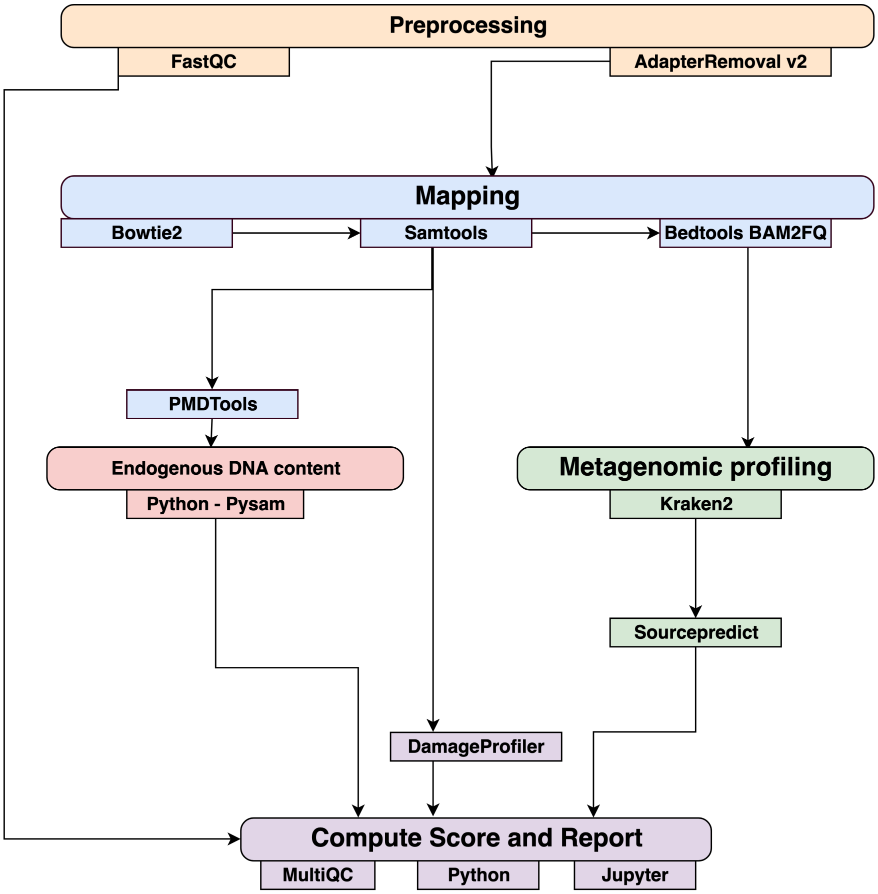


Figure 2. Workflow schematic of the coproID pipeline.
coproID consists of five steps: *Preprocessing* (orange), *Mapping* (blue), *Computing host DNA content for each metagenome* (red), *Metagenomic profiling* (green), and *Reporting* (violet). Individual programs (squared boxes) are colored by category(rounded boxes)

Afterwards, an host DNA ratio is computed for each source species such as:

$$NormalizedRatio(source\ species) = \frac{NormalizedHostDNA(source\ species)}{\sum NormalizedHost\ DNA\ (source\ species)} \tag{2}$$

where $\sum NormalizedHost\ DNA\ (source\ species)$ is the sum of all source species Normalized Host DNA.

Metagenomic profiling

Adapter clipped and trimmed reads are given as an input to Kraken 2 (Wood and Salzberg, 2014). Using the MiniKraken2_v2_8GB database (2019/04/23 version), Kraken 2 performs the taxonomic classification to output a taxon count per sample report file. All samples taxon count are pooled together in a taxon counts matrix with samples in columns, and taxons in rows. Next, Sourcepredict (Borry, 2019b) is used to predict the source based on each microbiome sample taxon composition. Using dimension reduction and K-Nearest Neighbors (KNN) machine learning trained with reference modern gut microbiomes samples (Table 1), Sourcepredict estimates a proportion $prop_{microbiome}(source\ species)$ of each potential source species, here Human or Dog, for each sample.

Reporting

For each filtered alignment file, the DNA damage patterns are estimated with DamageProfiler (Peltzer and Neukamm, 2019). The information from the host DNA content and the metagenomic profiling are gathered for each source in each sample such as:

$$proportion(source\ species) = NormalizedRatio(source\ species) \cdot prop_{microbiome}(source\ species)$$

Finally, a summary report is generated including the damage plots, a summary table of the coproID metrics, and the embedding of the samples in two dimensions by Sourcepredict. coproID is available on GitHub at the following address: **github.com/nf-core/coproid**.

RESULTS

We analyzed 21 archaeological samples with coproID v1.0 to estimate their source using both host DNA and microbiome composition.

Host DNA in reference gut microbiomes

Before analyzing the archaeological samples, we first tested whether there is a per-species difference in host DNA content in modern reference human and dog feces. With Anonymap, we computed the amount of host DNA in each reference gut microbiome (Table S1). We found that the median percentages of host DNA in NWHR, WHU, and Dog (Figure 3) are significantly different at $alpha = 0.05$ (Kruskal-Wallis H-test = 117.40, p value < 0.0001). We confirmed that there is a significant difference of median percentages of host DNA between dogs and NWHR, as well as dogs and WHU, with Mann-Whitney U tests (Table 3) and therefore corrected each sample by the mean percentage of gut host DNA found in each species, 1.24% for humans ($\mu_{NWHR} = 0.85$, $\sigma_{NWHR} = 2.33$, $\mu_{WHU} = 1.67$, $\sigma_{WHU} 0.81$), and 0.11% for dogs ($\sigma_{dog} = 0.16$) (equation 1, table S1). This information was used to correct for the amount of host DNA found in paleofeces.

Comparison	Mann–Whitney U test	p value
Dog vs NWHR	3327.0	< 0.0001
Dog vs WHU	41.0	< 0.0001
NWHR vs WHU	370.0	< 0.0001
Dog vs Human	3368.0	< 0.0001

Table 3. Statistical comparison of reference gut host DNA content. Mann–Whitney U test for independent observations . $H0$: the distributions of both populations are equal.

The effect of PMD filtering on host species prediction

Because aDNA accumulates damage over time (Briggs et al., 2007), we could use this characteristic to filter for reads carrying these specific damage patterns using PMDtools, and therefore reduce modern contamination in the dataset. We applied PMD filtering to our archaeological datasets, and for each, compared the predicted host source before and afterwards. The predicted host sources did not change after the DNA damage read filtering, but some became less certain (Figure 4). Most samples are confidently assigned to one of the two target species, however some samples previously categorized as humans now lie in the uncertainty zone. This suggests that PMDtools filtering lowered the modern human contamination which might have originated from sample excavation and manipulation.

The trade-off of PMDtools filtering is that it reduces the assignment power by lowering the number of reads available for host DNA based source prediction by only keeping PMD-bearing reads. This loss is greater for well-preserved samples, which may have relatively few damaged reads ($< 15\%$ of total). Ultimately, applying damage filtering can make it more difficult to categorize samples on the sole basis of host DNA content, but it also makes source assignments more reliable by removing modern contamination.

Source microbiome prediction of reference samples by Sourcepredict

To help resolve ambiguities related to the host aDNA present within a sample, we also investigated gut microbiome composition as an additional line of evidence to better predict paleofeces source. After performing taxonomic classification using Kraken2, we computed a sample pairwise distance matrix from the species counts. With the t-SNE dimension reduction method, we embedded this distance matrix in

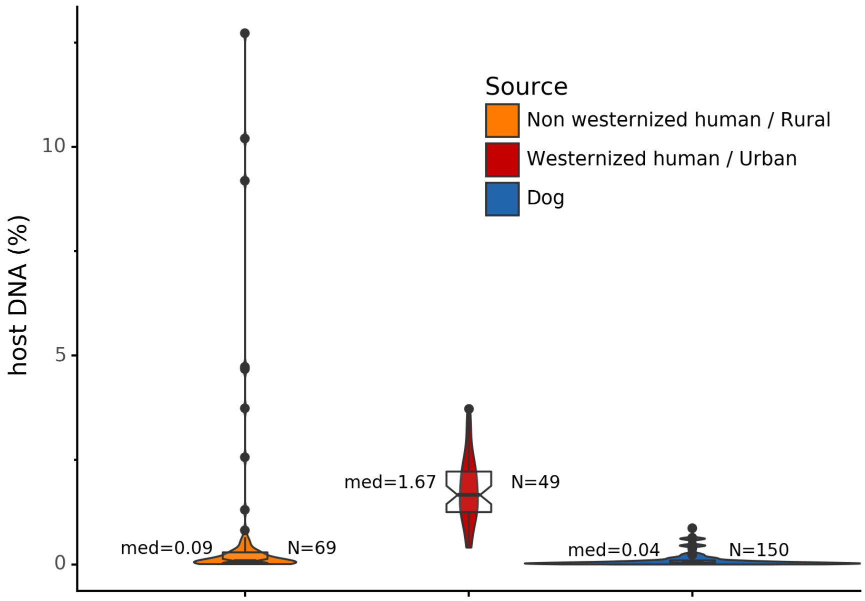


Figure 3. Gut microbiome host DNA content.
The median percentage of host DNA in the gut microbiome and the number of samples in each group are displayed besides each boxplot.

two dimensions to visualize the sample positions and sources (Figure 5a). We then used a KNN machine learning classifier on this low dimension embedding to predict the source of gut microbiome samples. This trained KNN model reached a test accuracy of 0.94 on previously unseen data (figure 5b).

Embedding of archaeological samples by Sourcepredict

We used this trained KNN model to predict the sources of the 20 paleofeces and coprolite archaeological samples, after embedding them in a two-dimensional space (Figure 6). Based on their microbiome composition data, Sourcepredict predicted 2 paleofeces samples as dogs, 8 paleofeces samples as human, 2 paleofeces samples and 4 archaeological sediments as soil, while the rest were predicted as unknown (Table S2).

coproID prediction

Combining both PMD-filtered host DNA information and microbiome composition, coproID was able to reliably categorize 7 of the 13 paleofeces samples, as 5 human paleofeces and 2 canine paleofeces, whereas all of the non-fecal archaeological sediments were flagged as unknown. (Figure 8). This confirms the original archaeological source hypothesis for five samples (ZSM005, ZSM025, ZSM027, ZSM028, ZSM031) and specifies or rejects the original archaeological source hypothesis for the two others (YRK001, AHP004). The 6 paleofeces samples not reliably identified by coproID have a conflicting source proportion estimation between host DNA and microbiome composition (Figure 7a and 7b and Table S3). Specifically, paleofeces AHP001, AHP002, and AHP003 show little predicted gut microbiome preservation, and thus have likely been altered by taphonomic (decomposition) processes. Paleofeces ZSM002, ZSM023, and ZSM029, by contrast, show good evidence of both host and microbiome preservation, but have conflicting source predictions based on host and microbiome evidence. Given that subsistence is associated with gut microbiome composition, this conflict may be related to insufficient gut microbiome datasets available for non-Westernized dog populations (Hagan et al., 2019).

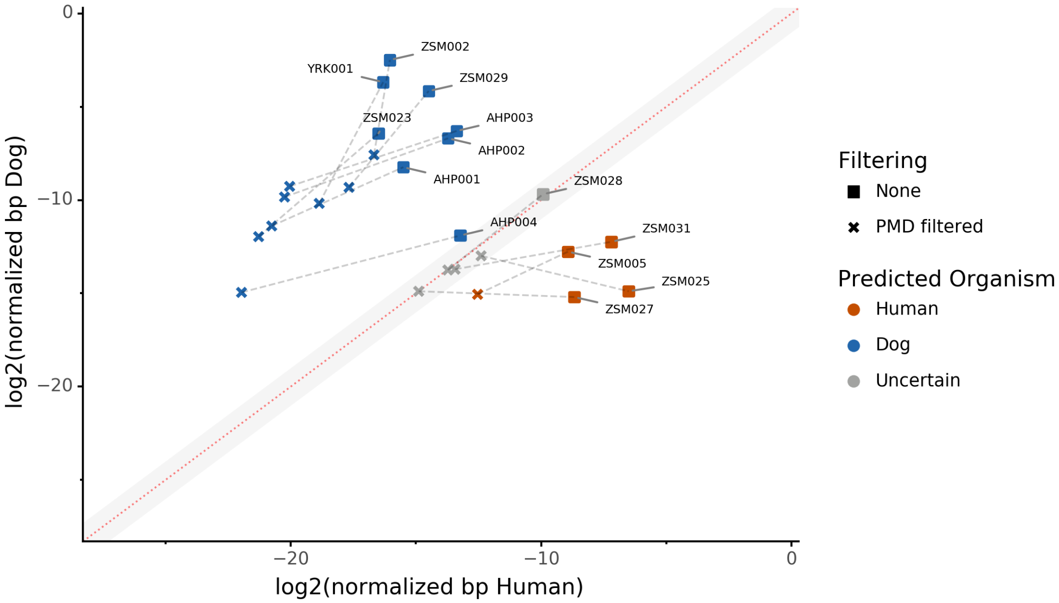
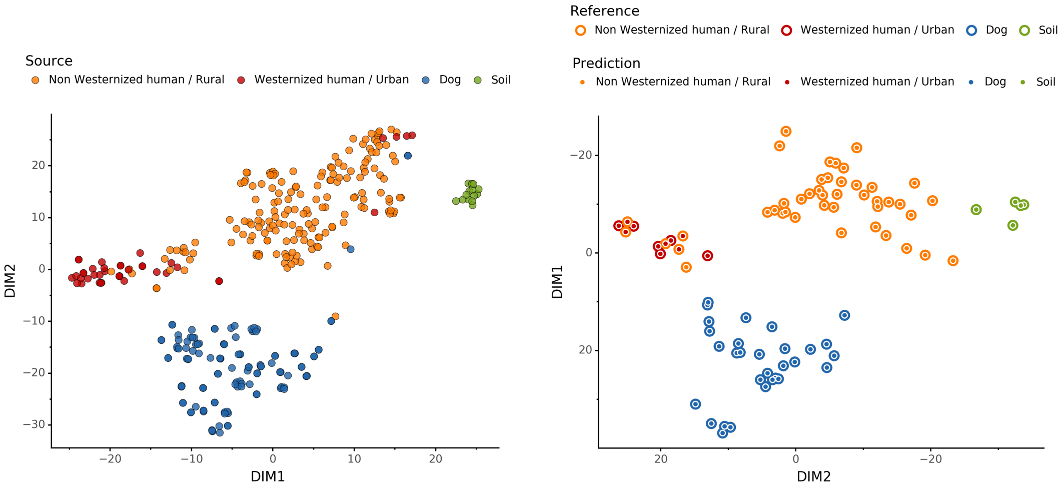


Figure 4. The effect of filtering for damaged reads using PMD.
The $\log_2$ of the human $NormalizedHostDNA$ is graphed against the $\log_2$ of the dog $NormalizedHostDNA$. Squares represent samples before filtering by PMD, whereas crosses represent samples after filtering by PMD. Dotted lines show the correspondence between samples. The red diagonal line marks the boundary between the two species, and the grey shaded area indicates a zone of species uncertainty ($\pm$ 1 $log2FC$) due to insufficient genetic information.



(a) Modern training samples

(b) Modern test samples

Figure 5. Embedding of reference modern gut microbiomes.
(a) t-SNE embedding of the species composition based on sample pairwise Weighted Unifrac distances for training modern gut microbiomes training samples. Samples are colored by their actual source. **(b)** t-SNE embedding of the species composition based on sample pairwise Weighted Unifrac distances for source prediction of modern test samples. The outer circle color is the actual source of a sample, while the inner circle color is the predicted sample source by Sourcepredict.

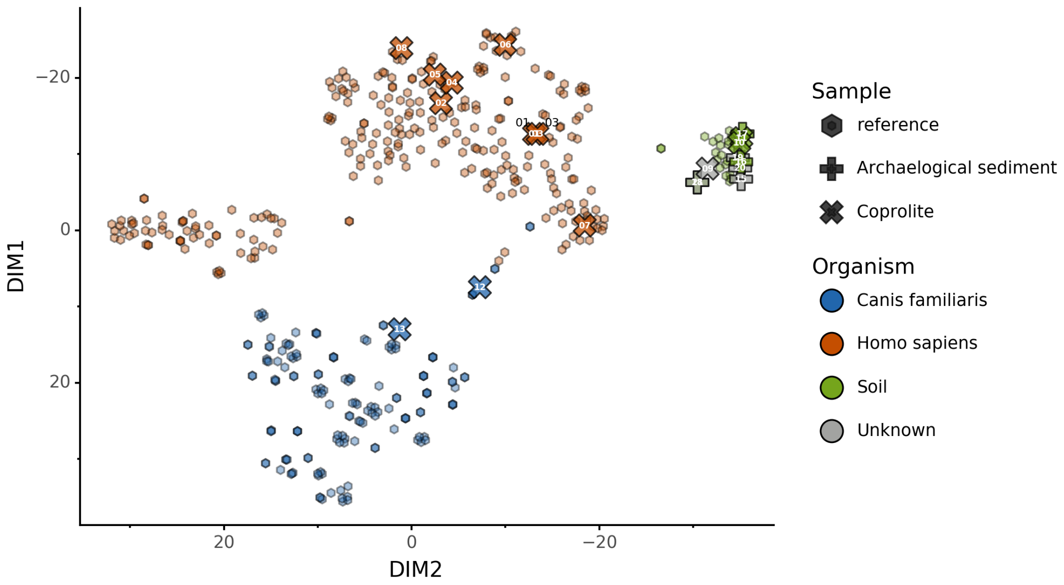
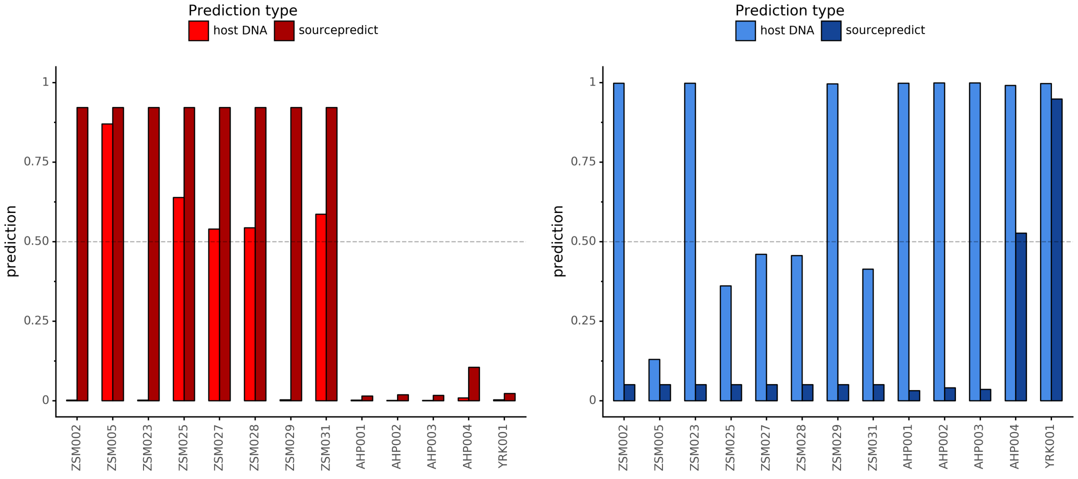


Figure 6. Prediction of archaeological samples sources and t-SNE embedding by Sourcepredict. t-SNE embedding of archaeological (crosses) and modern (hexagons) samples. The color of the modern samples is based on their actual source while the color of the archaeological samples is based on their predicted source by Sourcepredict. Archaeological sample are labelled with their *Plot ID* (Table 2).



(a) coproID human prediction

(b) coproID canine prediction

Figure 7. Host DNA and Sourcepredict source prediction for paleofeces samples. The vertical bar represents the predicted proportion by host DNA (lighter fill) or by Sourcepredict (darker fill). The horizontal dashed line represents the confidence threshold to assign a source to a sample.

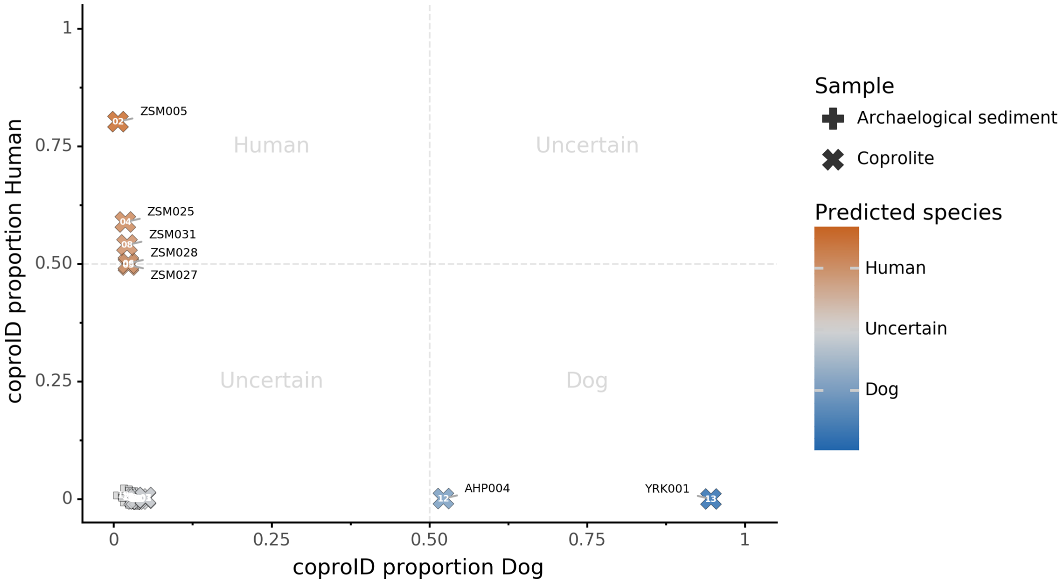


Figure 8. **coproID source prediction**.
Predicted human proportion graphed versus predicted canine proportion. Samples are colored by their predicted sources proportions. Samples with a low canine and human proportion are not annotated.

DISCUSSION

Paleofeces are the preserved remains of human or animal feces, and although they typically only preserve under highly particular conditions, they are nevertheless widely reported in the paleontological and archaeological records and include specimens ranging in age from the Paleozoic era (Dentzien-Dias et al., 2013) to the last few centuries. Paleofeces can provide unprecedented insights into animal health and diet, parasite biology and evolution, and the changing ecology and evolution of the gut microbiome. However, because many paleofeces lack distinctive morphological features, determining the host origin of a paleofeces can be a difficult problem (Poinar et al., 2009). In particular, distinguishing human and canine paleofeces can be challenging because they are often similar in size and shape, they tend to co-occur at archaeological sites and in midden deposits, and humans and domesticated dogs tend to eat similar diets (Guiry, 2012). We developed coproID to aid in identifying the source organism of archaeological paleofeces and coprolites by applying a combined approach relying on both ancient host DNA content and gut microbiome composition.

coproID addresses several shortcomings of previous methods. First, we have included a DNA damage-filtering step that allows for the removal of potentially contaminating modern human DNA, which may otherwise skew host species assignment. We have additionally measured and accounted for significant differences in the mean proportion of host DNA found in dog and human feces, and we also accounted for differences in host genome size between humans and dogs when making quantitative comparisons of host DNA. Then, because animal DNA recovered from paleofeces may contain a mixture of host and dietary DNA, we also utilize gut microbiome compositional data to estimate host source. We show that humans and dogs have distinct gut microbiome compositions, and that their feces can be accurately distinguished from each other and from non-feces using a machine learning classifier after data dimensionality reduction. Taken together, these approaches allow a robust determination of paleofeces and coprolite host source, that takes into account both modern contamination, microbiome composition, and postmortem degradation.

In applying coproID to a set of 20 archaeological samples of known and/or suspected origin, all 7 non-fecal sediment samples were accurately classified as ”uncertain” and were grouped with soil by Sourcepredict. For the 13 paleofeces and coprolites under study, 7 exhibited matching host and microbiome source assignments and were confidently classified as either human (n=5) or canine (n=2). Importantly, one of the samples confidently identified as canine was YRK001, a paleofeces that had been recovered from an archaeological chamber pot in the United Kingdom, but which showed an unusual diversity of parasites inconsistent with human feces, and therefore posed issues in host assignation.

For the remaining six unidentified paleofeces, three exhibited poor microbiome preservation and were classified as ”uncertain”, while the other three were well-preserved but yielded conflicting host DNA and microbiome assignments. These three samples, ZSM002, Z023, and ZSM029, all from prehistoric Mexico, all contain high levels of canine DNA, but have gut microbiome profiles within the range of NWHR humans. Classified as ”uncertain”, there are two possible explanations for these samples. First, these feces could have originated from a human who consumed a recent meal of canine meat. Dogs were consumed in ancient Mesoamerica (Clutton-Brock and Hammond, 1994; Santley and Rose, 1979; Rosenswig, 2007; Wing, 1978), but further research on the expected proportion of dietary DNA in human feces is needed to determine whether this is a plausible explanation for the very high amounts of canine DNA (and negligible amounts of human DNA) observed.

Alternatively, these feces could have originated from a canine whose microbiome composition is shifted relative to that of the reference metagenomes used in our training set. It is now well-established that subsistence mode strongly influences gut microbiome composition in humans Obregon-Tito et al. (2015), with NWHR and WHU human populations largely exhibiting distinct gut microbiome structure, as seen in (Figure 5a. To date, no gut microbiome data is available from non-Westernized dogs, and all reference dog metagenome data included as training data for coproID originated from a single study of labrador retrievers and beagles Coelho et al. (2018). Future studies of non-Westernized rural dogs are needed to establish the full range of gut microbial diversity in dogs and to more accurately model dog gut microbiome diversity in the past. Given that all confirmed human paleofeces in this study falls within the NWHR cluster (Figure 6), we anticipate that our ability to accurately classify dog paleofeces and coprolites as canine (as opposed to ”uncertain”) will improve with the future addition of non-Westernized rural dog metagenomic data.

CONCLUSIONS

We developed an open-source, documented, tested, scalable, and reproducible method to perform the identification of archaeological paleofeces and coprolite source. By leveraging the information from host DNA and microbiome composition, we were able to identify and/or confirm the source of newly sequenced paleofeces. We demonstrated that coproID can provide useful assistance to archaeologists in identifying authentic paleofeces and inferring their host. Future work on dog gut microbiome diversity, especially among rural, non-Westernized dogs, may help improve the tool's sensitivity even further.

ACKNOWLEDGMENTS

We thank David Petts, Zdeněk Tvrdý, Susanne Stegmann-Rajtár, and Zuzana Rajtarova for contributing archaeological samples to this study. We thank the Guildford Museum (Guildford Borough Council Heritage Service) and Catriona Wilson for allowing us to analyze the chamber pot paleofeces sample from Surrey, UK. The sample from Derragh, Ireland was excavated by Discovery Programme, an all-Ireland public center of archaeological research supported by the Heritage Council, during field work in 2003 to 2005 as part of the Lake Settlement Project. This work was supported by the US National Institutes of Health R01GM089886 (to C.W. and C.M.L.), the Deutsche Forschungsgemeinschaft EXC 2051 #390713860 (to C.W.), and the Max Planck Society. Author contributions were as follows: M.B. and C.W. designed the research. B.C., M.C.W., D.J., C.A.H., and R.W.H. performed the laboratory experiments. M.B., C.J., M.C.W, and T.H. analyzed the data. M.B. developed the bioinformatics tools and pipelines. C.W., K.R., K.B., L.G.F., A.P., A.K., W.T.J.K, R.P., I.S., D.S.G., J.Y., T.S.K, N.M., H.C., and C.M.L. provided materials and resources. C.W., A.He., and A.Hü. supervised the research. M.B. wrote the article, with input from C.W., A.Hü., KR and the other co-authors.

DATA AND CODE AVAILABILITY

Genetic data are available in the European Nucleotide Archive (ERA) under the accessions PRJEB33577 and PRJEB35362. The code for the analysis is available at github.com/maxibor/coproid-article.

REFERENCES

Andrews, S. et al. (2010). Fastqc: a quality control tool for high throughput sequence data.

Bon, C., Berthonaud, V., Maksud, F., Labadie, K., Poulain, J., Artiguenave, F., Wincker, P., Aury, J.-M., and Elalouf, J.-M. (2012). Coprolites as a source of information on the genome and diet of the cave hyena. *Proceedings of the Royal Society B: Biological Sciences*, 279(1739):2825–2830.

Borry, M. (2019a). maxibor/anonymap: Anonymap v1.0.

Borry, M. (2019b). Sourcepredict: Prediction of metagenomic sample sources using dimension reduction followed by machine learning classification. *Journal of Open Source Software*, 4(41):1540.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, 104(37):14616–14621.

Brito, I. L., Gurry, T., Zhao, S., Huang, K., Young, S. K., Shea, T. P., Naisilisili, W., Jenkins, A. P., Jupiter, S. D., Gevers, D., and Alm, E. J. (2019). Transmission of human-associated microbiota along family and social networks. *Nature Microbiology*, page 1.

Butler, J. and Du Toit, J. (2002). Diet of free-ranging domestic dogs (canis familiaris) in rural zimbabwe: implications for wild scavengers on the periphery of wildlife reserves. In *Animal Conservation forum*, volume 5, pages 29–37. Cambridge University Press.

Clutton-Brock, J. and Hammond, N. (1994). Hot dogs: comestible canids in preclassic maya culture at cuello, belize. *Journal of Archaeological Science*, 21(6):819–826.

Coelho, L. P., Kultima, J. R., Costea, P. I., Fournier, C., Pan, Y., Czarnecki-Maulden, G., Hayward, M. R., Forslund, S. K., Schmidt, T. S. B., Descombes, P., Jackson, J. R., Li, Q., and Bork, P. (2018). Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome*, 6(1):72.

CSIR, C. i. o. m. and aromatic plants (2016). Chrysopogon zizanioides (ID 322597) - BioProject - NCBI.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., et al. (2013). Complete mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort dna fragments. *Proceedings of the National Academy of Sciences*, 110(39):15758–15763.

Davenport, E. R., Sanders, J. G., Song, S. J., Amato, K. R., Clark, A. G., and Knight, R. (2017). The human microbiome in evolution. *BMC biology*, 15(1):127.

Dentzien-Dias, P. C., Poinar Jr, G., de Figueiredo, A. E. Q., Pacheco, A. C. L., Horn, B. L., and Schultz, C. L. (2013). Tapeworm eggs in a 270 million-year-old shark coprolite. *PLoS One*, 8(1):e55007.

Dhakan, D. B., Maji, A., Sharma, A. K., Saxena, R., Pulikkan, J., Grace, T., Gomez, A., Scaria, J., Amato, K. R., and Sharma, V. K. (2019). The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *GigaScience*, 8(3).

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature biotechnology*, 35(4):316.

Ewels, P., Peltzer, A., Fillinger, S., Alneberg, J., Patel, H., Wilm, A., Garcia, M., Di Tommaso, P., and Nahnsen, S. (2019). nf-core: Community curated bioinformatics pipelines. *bioRxiv*, page 610741.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., Owens, S., Gilbert, J. A., Wall, D. H., and Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, 109(52):21390–21395.

Frantz, L. A., Mullin, V. E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., Linderholm, A., Mattiangeli, V., Teasdale, M. D., Dimopoulos, E. A., et al. (2016). Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290):1228–1231.

Gilbert, M. T. P., Jenkins, D. L., Götherstrom, A., Naveran, N., Sanchez, J. J., Hofreiter, M., Thomsen, P. F., Binladen, J., Higham, T. F., Yohe, R. M., et al. (2008). Dna from pre-clovis human coprolites in oregon, north america. *Science*, 320(5877):786–789.

Guiry, E. J. (2012). Dogs as analogs in stable isotope-based human paleodietary reconstructions: a review and considerations for future use. *Journal of Archaeological Method and Theory*, 19(3):351–376.

Hagan, R. W., Hofman, C. A., Hübner, A., Reinhard, K., Schnorr, S., Lewis, C. M., Sankaranarayanan, K., and Warinner, C. G. (2019). Comparison of extraction methods for recovering ancient microbial dna from paleofeces. *American Journal of Physical Anthropology*.

Hofreiter, M., Poinar, H. N., Spaulding, W. G., Bauer, K., Martin, P. S., Possnert, G., and Pääbo, S. (2000). A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology*, 9(12):1975–1984.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., et al. (2012). Structure, function and diversity of the healthy human microbiome. *nature*, 486(7402):207.

Jiménez, F. A., Gardner, S. L., Araújo, A., Fugassa, M., Brooks, R. H., Racz, E., and Reinhard, K. J. (2012). Zoonotic and human parasites of inhabitants of cueva de los muertos chiquitos, rio zape valley, durango, mexico. *Journal of Parasitology*, 98(2):304–310.

Kho, Z. Y. and Lal, S. K. (2018). The human gut microbiome–a potential controller of wellness and disease. *Frontiers in microbiology*, 9.

Kirch, P. and O'Day, S. J. (2003). New archaeological insights into food and status: a case study from pre-contact hawaii. *World Archaeology*, 34(3):484–497.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R., and Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, 8(9):761–763.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R., et al. (2008). Evolution of mammals and their gut microbes. *Science*, 320(5883):1647–1651.

Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Yates, J. A. F., Salazar-García, D. C., et al. (2018). Differential preservation of endogenous human and microbial dna in dental calculus and dentin. *Scientific reports*, 8(1):9822.

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6):pdb–prot5448.

Obregon-Tito, A. J., Tito, R. Y., Metcalf, J., Sankaranarayanan, K., Clemente, J. C., Ursell, L. K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P. M., Spicer, P., Lawson, P., Marin-Reyes, L., Trujillo-Villarroel, O., Foster, M., Guija-Poma, E., Troncoso-Corzo, L., Warinner, C., Ozga, A. T., and Lewis, C. M. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Communications*, 6:6505.

Orellana, L. H., Chee-Sanford, J. C., Sanford, R. A., Löffler, F. E., and Konstantinidis, K. T. (2018). Year-Round Shotgun Metagenomes Reveal Stable Microbial Communities in Agricultural Soils and Novel Ammonia Oxidizers Responding to Fertilization. *Applied and Environmental Microbiology*, 84(2):e01646–17.

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., and Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, 176(3):649–662.e20.

Peltzer, A. and Neukamm, J. (2019). Integrative-Transcriptomics/DamageProfiler: DamageProfiler v0.4.7.

Podberscek, A. L. (2009). Good to pet and eat: The keeping and consuming of dogs and cats in south korea. *Journal of Social Issues*, 65(3):615–632.

Poinar, H., Fiedel, S., King, C. E., Devault, A. M., Bos, K., Kuch, M., and Debruyne, R. (2009). Comment on “dna from pre-clovis human coprolites in oregon, north america”. *Science*, 325(5937):148–148.

Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershed, R. P., Possnert, G., and Pääbo, S. (1998). Molecular coproscopy: dung and diet of the extinct ground sloth nothrotheriops shastensis. *Science*, 281(5375):402–406.

Poinar, H. N., Kuch, M., Sobolik, K. D., Barnes, I., Stankiewicz, A. B., Kuder, T., Spaulding, W. G., Bryant, V. M., Cooper, A., and Pääbo, S. (2001). A molecular analysis of dietary diversity for three archaic native americans. *Proceedings of the National Academy of Sciences*, 98(8):4317–4322.

pysam developers (2018). Pysam: a python module for reading and manipulating files in the sam/bam format.

Rampelli, S., Schnorr, S., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., Brigidi, P., Crittenden, A., Henry, A., and Candela, M. (2015). Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Current Biology*, 25(13):1682–1693.

Rosenswig, R. M. (2007). Beyond identifying elites: Feasting as a means to understand early middle formative society on the pacific coast of mexico. *Journal of Anthropological Archaeology*, 26(1):1–27.

Santley, R. S. and Rose, E. K. (1979). Diet, nutrition and population dynamics in the basin of mexico. *World Archaeology*, 11(2):185–207.

Schubert, M., Lindgreen, S., and Orlando, L. (2016). Adapterremoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes*, 9(1):88.

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in plant science*, 5:209.

Shenhav, L., Thompson, M., Joseph, T. A., Briscoe, L., Furman, O., Bogumil, D., Mizrahi, I., Pe’er, I., and Halperin, E. (2019). FEAST: fast expectation-maximization for microbial source tracking. *Nature Methods*, page 1.

Sistiaga, A., Mallol, C., Galván, B., and Summons, R. E. (2014). The neanderthal meal: a new perspective using faecal biomarkers. *PloS one*, 9(6):e101045.

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences*, 111(6):2229–2234.

The Human Microbiome Project Consortium, Huttenhower, C., Gevers, D., Knight, Rob, W. O., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214.

Tito, R. Y., Knights, D., Metcalf, J., Obregon-Tito, A. J., Cleeland, L., Najar, F., Roe, B., Reinhard, K., Sobolik, K., Belknap, S., Foster, M., Spicer, P., Knight, R., and Lewis, C. M. (2012). Insights from Characterizing Extinct Human Gut Microbiomes. *PLoS ONE*, 7(12):e51146.

Tito, R. Y., Macmil, S., Wiley, G., Najar, F., Cleeland, L., Qu, C., Wang, P., Romagne, F., Leonard, S., Ruiz, A. J., et al. (2008). Phylotyping and functional analysis of two ancient human microbiomes. *PLoS One*, 3(11):e3703.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., Burbano, H. A., Orlando, L., and Krause, J. (2017). A robust framework for microbial archaeology. *Annual review of genomics and* *human genetics*, 18:321–356.

Warinner, C. and Lewis Jr, C. M. (2015). Microbiome and health in past and present human populations. *American Anthropologist*, 117(4):740–741.

Warinner, C., Speller, C., Collins, M. J., and Lewis Jr, C. M. (2015). Ancient human microbiomes. *Journal of human evolution*, 79:125–136.

Wibowo, M. C., Yang, Z., Tierney, B. T., Luber, J. M., Barajas-Olmos, F., Cecilia, C.-C., Humberto, G.-O., Martinez-Hernandez, A., Zimmerman, S., Smiley, F. E., Ballal, S. A., Reinhard, K., Russ, J., Orozco, L., Snow, M., LeBlanc, S., and Kostic, A. D. (2019). Reconstruction of ancient microbial genomes from the human gut - in review. in review.

Wing, E. S. (1978). *Use of dogs for food: An adaptation to the coastal environment*. Elsevier.

Wood, D. E. and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, 15(3):R46.

Wood, J. R., Crown, A., Cole, T. L., and Wilmshurst, J. M. (2016). Microscopic and ancient dna profiling of polynesian dog (kurī) coprolites from northern new zealand. *Journal of Archaeological Science: Reports*, 6:496–505.