[Supplementary material]

A Black Death mass grave at Thornton Abbey: the discovery and examination of a fourteenth-century rural catastrophe

Hugh Willmott[1,*], Peter Townend[2], Diana Mahoney Swales[3], Hendrik Poinar[4], Katherine Eaton[4] & Jennifer Klunk[4]

[1] *Department of Archaeology, University of Sheffield, UK*

[2] *Network Archaeology, Lincoln, UK*

[3] *Centre for Anatomy & Human Identification, University of Dundee, UK*

[4] *McMaster Ancient DNA Centre, McMaster University, Canada*

[*] *Author for correspondence (Email: h.willmott@sheffield.ac.uk)*

Ancient DNA laboratory protocols

DNA extraction

Ancient DNA laboratory work was performed in dedicated clean-room facilities at the McMaster Ancient DNA Centre (Hamilton, ON). One root of each individual's molar was sectioned, manually crushed, and divided into two subsamples (50–100mg each). Reagent blanks were introduced as negative controls to monitor DNA contamination in subsequent steps (referred to as "Extraction Blank Controls" or "EBC"). Demineralisation and digestion were performed as previously described (Schwarz *et al*. 2009) and DNA extraction was conducted following the method outlined in Dabney *et al*. (2013).

Plague PCR screening

An initial plague-screening PCR was performed in duplicate on 1:10 extract dilutions following the *pla* assay in Wagner *et al*. (2014). In brief, the PCR primers used in this study target the 3'UTR of the *pla* gene which has reduced sequence similarity in non-*Yersinia* species such as *Escherichia coli* and *Citrobacter koseri*. The forward primer used is thus far known to be identical to *Yersinia pestis* only. Only Sk.36 and Sk.45 demonstrated amplification with melt peaks within 0.5 degrees of the standard (Table S1). A second round of *pla* PCR was performed for each of Sk.36 and Sk.45 using the 1:10 dilution in duplicate and the original concentration in duplicate.

<TABLE S1>

Sequencing library preparation

The extracted DNA of Sk.36 and Sk.45 was converted into Illumina sequencing libraries using a modified protocol (Kircher *et al*. 2012) and quantified using an Illumina library qPCR assay. Library preparation was performed in duplicate for each extract (hereafter designated S1 and S2) on separate days to monitor batch effects. A new negative control was introduced at this stage (referred to as the "Library Blank Control" or "LBC") for each of S1 and S2. Sample Sk.45 S1 appeared to be affected by ethanol contamination during library preparation as there was excess volume produced after the second column purification step and it performed poorly in the total-library quant. As a result, Sk.45 S1 did not meet sequencing concentration requirements and was re-amplified for 10 cycles following the previous qPCR assay. Sample libraries were pooled at equimolar concentrations while negative controls were sequenced at maximum volume input to maximize detection of contaminant organisms. Paired-end sequencing was performed on an Illumina HiSeq 1500 platform at the Farncombe Metagenomics Facility (Hamilton, ON).

Enrichment

In-solution enrichment for the pan-genome of *Y. pestis* was performed with the probes designed in Wagner *et al*. (2014) and using the myBaits v3 protocol (Arbor Biosciences, Ann Arbor, MI). Sk.36 was enriched using the library generated from S2 (as the S1 library volume was exhausted) and alongside another extraction blank control (designated EBCE afterwards). The following modifications were incorporated to improve recovery of degraded and divergent DNA sequences: 5uL library input, 100ng bait concentration, hybridisation at 55°C, 16–24h hybridisation capture, and two rounds of enrichment. The enriched libraries were quantified using an Illumina library quantification qPCR assay and pooled at maximum input volume (13uL) due to low concentration. Following pooling, libraries were size-selected on agarose gel to retain 150–500bp fragments which corresponds to molecule lengths of approximately 15–365bp without the adapter sequences. Paired-end sequencing was performed on an Illumina HiSeq 1500 platform at the Farncombe Metagenomics Facility (Hamilton, ON).

Computational analysis

Pre-processing

Sequence data were pre-processed following typical ancient DNA methods including trimming and merging (Renaud *et al*. 2014), adapter-filtering, length-filtering, and duplicate removal. Only paired-end molecules that could be successfully merged were used for metagenomic analysis, and a minimum length of 35bp was enforced as this was required by downstream programs. Sk.45 S1 sequenced poorly as expected due to the ethanol contamination as only 1546 molecules were sequenced as opposed to the remaining libraries which produced 2–5 million molecules (Table S2).

<TABLE S2>

Taxonomic composition

Sample preservation was assessed using a metagenomic approach to identify surviving host molecules (*Homo sapiens*) as well as oral microbiome bacteria and environmental taxa. Pre-processed libraries for Sk.36 S1, Sk.36 S2, and Sk.45 S2 were down-sampled to 250 000 molecules to improve computation speed. All other samples produced less reads and were analysed using their full pre-processed read set.

Molecules were first classified using the program Kraken2 (Wood & Salzberg 2014) against the standard database of Bacteria, Viruses, Archaea, Human, and Known Vectors and filtered for a minimum confidence level of 0.2. Successfully classified molecules were piped into the second classification program, blasn (Altschul *et al*. 1990) running against the nt database (last updated in June 2018), retaining the first 100 matches for each molecule and filtering for an e-value <=0.00001. Following this step, only top hits for each molecules (tied e-value) were retained, and the taxonomic classifications were summarized using a lowest-common-ancestor (LCA) approach as implemented in KronaTools (Ondov *et al*. 2011). From this output, a consensus file was generated to report on species and species complex classifications that were concordant between the two programs per molecule.

All sequenced samples were clustered based on their concordant species classification counts transformed into proportion data. Bray-Curtis dissimilarity was used to measure profile differences and Ward's method was used for hierarchical clustering (Suzuki & Shimodaira 2006). Multi-scale bootstrap resampling was performed using 1000 bootstraps and clusters with an AU p-value of >=0.95 were marked. Note that while the p-value significance indicates these clusters are likely not due to sampling error, without sample replicates and data normalisation these clusters cannot be interpreted for significance beyond preliminary

exploratory analysis.

To assist in interpreting clusters, each sample's taxonomic species profile was depicted as stacked bar charts (Wickham 2009). As there were a large number of species identified across the study (n=1015), species were instead summarised and summed at the phylum level and the family level separately to provide a more condensed visualisation. At the family level, taxa present at less than 1 per cent were combined into an "Other" category. Hierarchical pie charts of the concordant species classifications were created using KronaTools to visualise the effect *Y. pestis* genome enrichment had upon the taxonomic profile for sample Sk.36

Plague identification

To identify plague-specific (*Y. pestis*) molecules in the high-depth libraries Sk.36 S1, Sk.36 S2 and Sk.45 S2, all pre-processed molecules were first classified with Kraken2 using the parameters stated previously. Molecules assigned within the genus *Yersinia* were subsequently classified by blastn. For all other samples and controls (low sequencing depth), all pre-processed molecules were used without the bottleneck of *Yersinia* selection in order to investigate potentially confounding taxa. All pre-processed molecules were also used as input for marker gene analysis with MetaPhlAn2 to detect diagnostic regions using the very-sensitive local alignment mode (Truong *et al*. 2015).

Genome enrichment data for sample Sk.36 was mapped to the *Y. pestis* reference genome CO92 (RefSeq Assembly: ASM906v1) including both merged and unmerged molecules. BWA was run with the following parameters modified for ancient DNA alignment: seeding disabled, a two gap open maximum, and an edit distance maximum of 0.01. Aligned molecules were filtered to retain only those that were mapped and properly paired (https://github.com/grenaud/libbam) and to remove coordinate-based PCR duplicates (https://bitbucket.org/ustenzel/biohazard). Alignments were also filtered to remove multi-mapped molecules and to retain a minimum mapping quality of 30. Finally, ancient DNA damage patterns (fragment length and cytosine deamination) were assessed using mapDamage (Jónsson *et al*. 2013).

Results

PCR plague screening

Of the 16 extracted samples, only Sk.36 and Sk.45 amplified using the *pla* assay (Table S1).

All 6/6 replicates of Sk.36 amplified with the expected product as indicated by being within 0.5 degrees of the standard in melt peak analysis. Only 1/6 replicates of Sk.45 amplified with the expected product. Sk.36 was tentatively classified as a “Strong” plague candidate while Sk.45 was classified as a “Weak/False-Positive” plague candidate.

Shotgun sequencing

Taxonomic species profile:

The species-profile clustering (Figures S1 and S2) suggest three sample groups:

1) Ancient shotgun samples:

Shotgun Sequencing of the ancient samples Sk36S1/S2 and Sk45S2.

2) Extraction blank controls

Extraction blank controls and the ethanol-inhibited replicate Sk45S1.

3) Enrichment and library blank

Y. pestis enrichments Sk36E, EBCE, and the library blank control LBCS2.

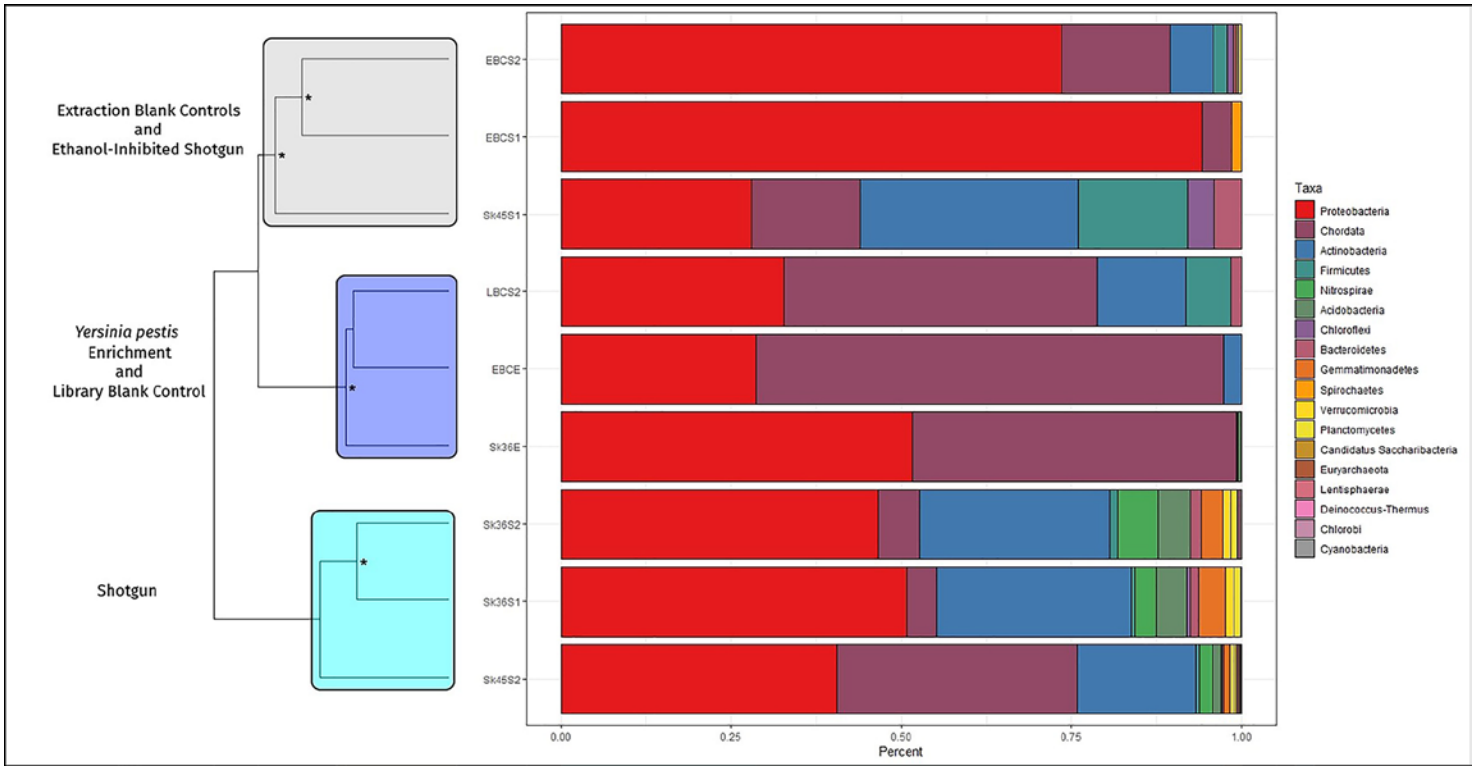

Figure S1. Hierarchical clustering of concordant species classifications with phylum proportions for visualisation.

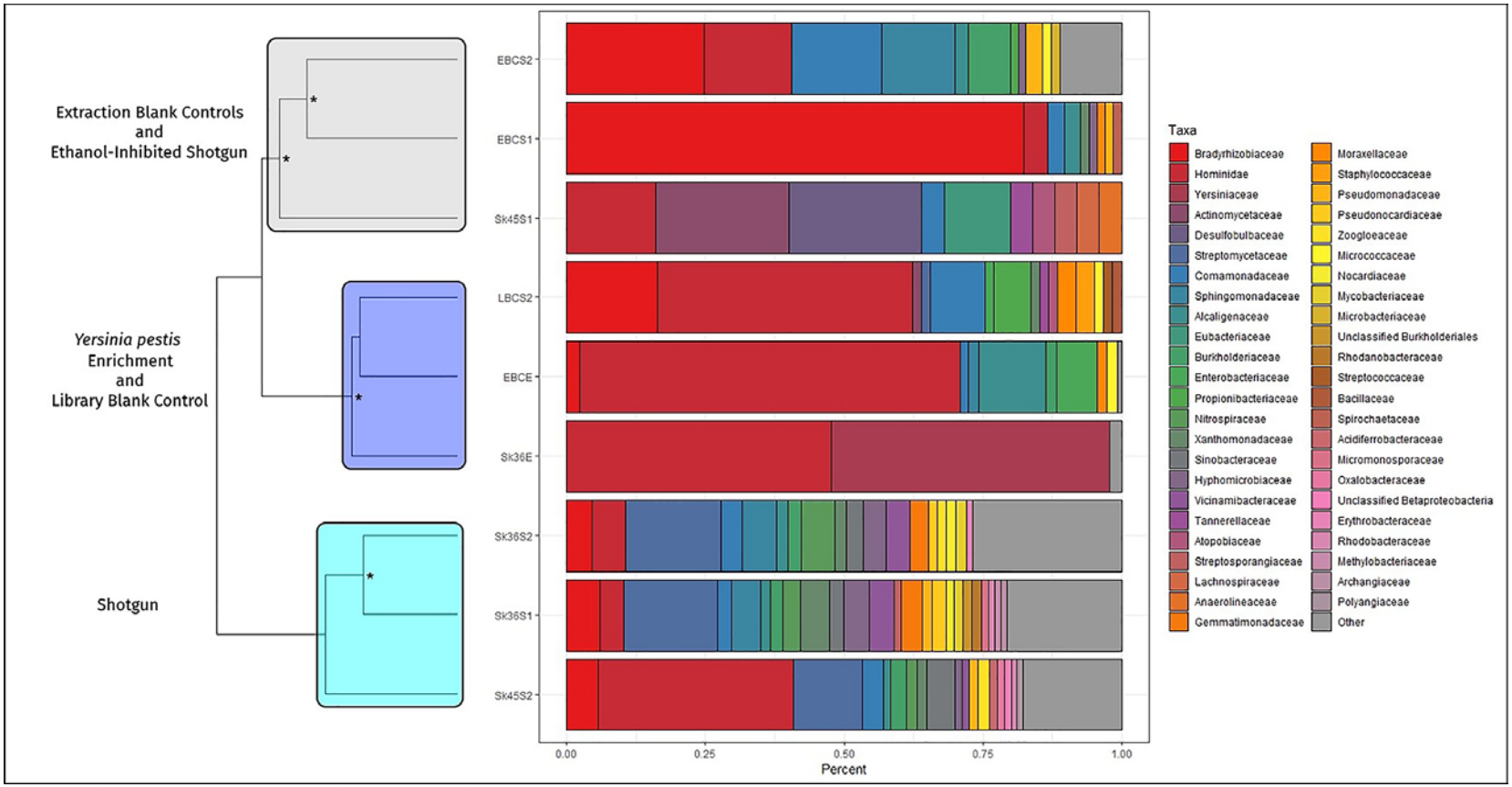

Figure S2. Hierarchical clustering of concordant species classifications with family proportions for visualisation.

The clustering pattern intuitively follows expectations that shotgun sequencing of samples ("unbiased") should be distinct from enriched samples (deliberate "bias") which should also be distinct from extraction blank controls. However, the third cluster (enrichment and the library blank) was unexpected but is suspected to be driven by the greater proportion of human DNA in these profiles.

An identical topology is produced when clustering at the genus and family level although all AU p-values drop below 0.95 except for the replicates Sk.36 S1 and Sk.36 S2. Clustering at the phylum level moves Sk.45 S2 into the enrichment cluster (#3), possibly due to the combination of moderate Proteobacteria/high Chordata while Sk. 45 S1 is moved into the shotgun cluster (#1) possibly because of low Proteobacteria/high Actinobacteria.

Library prep replicates of Sk.36 (S1 and S2) cluster together and appear to have very similar taxonomic compositions. Aside from the documented differential treatment of Sk.45 S1, there is no strong evidence for unforeseen batch effects driving compositional differences.

DNA Preservation

Human DNA

Sk. 36 and Sk. 45 appear to be in a state of poor molecular preservation as the proportion of *Homo sapiens* DNA is extremely low at <1 per cent (Table S3). In fact, in all cases, the negative DNA controls have a greater proportion of identified *Homo sapiens* DNA. However, enough molecules were retrieved to assess terminal aDNA damage patterns and all ancient samples (except the inhibited Sk.45 S1) demonstrate the expected inflation of C->T transitions within the 5-bp terminal ends. The library blank (LBCS2) also appears to contain degraded human DNA which may indicate sporadic cross-contamination from ancient samples to negative controls. While this cross-contamination does not appear to be ubiquitous across controls, this finding necessitates additional stringency in interpreting the low-abundance pathogen detection results and a careful consideration of negative control taxa.

<TABLE S3>

Oral Microbiome DNA

Neither Sk.36 (S1 or S2) nor Sk.45 (S2) have taxa associated with the oral microbiome as their dominant species (Table S5), which might be anticipated from sampling an archaeological tooth. Instead, the negative DNA clustered samples all contain an assortment of taxa that are

known to be associated with human microbiomes (Table S6). The phenomenon of reagent kit contamination with oral microbes has recently been published on (Weyrich *et al*. 2019) and thus these results are in agreement with current efforts to characterise background laboratory microbes.

Instead of a profile indicative of the oral microbiome, the profiles of Sk.36 and Sk.45 predominantly reflect a high abundance of environmental bacteria (soil, water, plant-associated). The shotgun sequenced samples share approximately 10% of species with the extraction blanks (Figure S3) however this proportion is highly influenced by sequencing depth variation and choice of subsampling level. When comparing the top ten most abundant species in each sample, only two species are shared between the ancient samples Sk.36/Sk.45 and the Extraction Blank Controls: *Homo sapiens* and a *Rhodoplanes* bacterium. Laboratory contamination is therefore unlikely to drive the signatures of high abundance species in this study, but low abundance taxa will require additional criteria for authentication.

Overall, the taxonomic profiles of Sk.36 and Sk.45 indicate poor preservation of endogenous (host) DNA and reflect substantial DNA exchange with post-mortem environments (burial context, museum curation, laboratory surfaces, etc.)

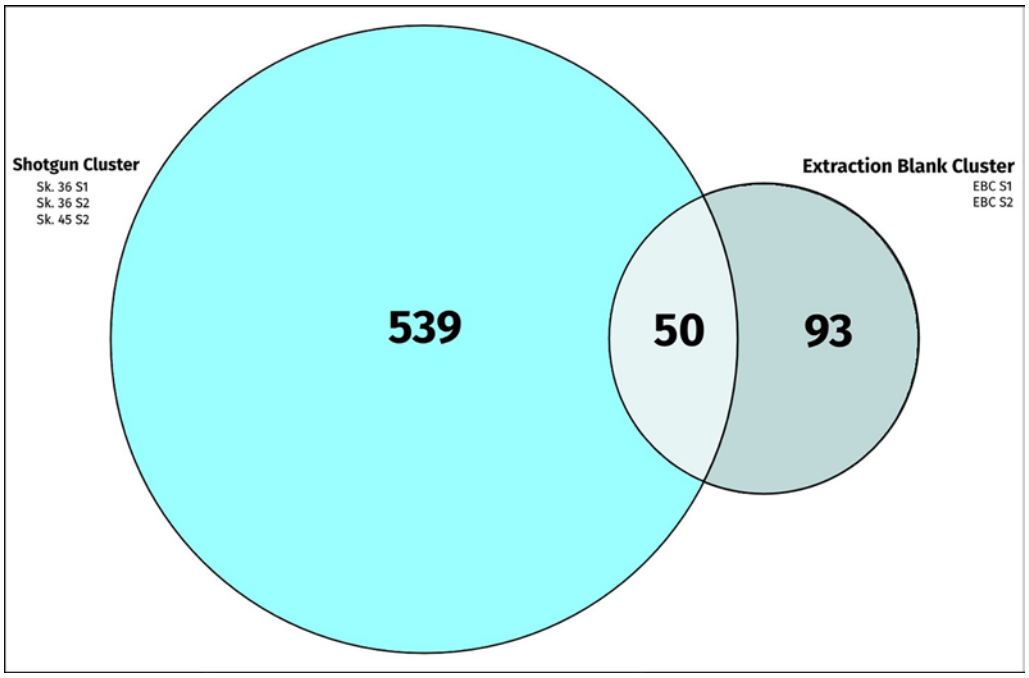

Figure S3. Unique and shared species between the ancient shotgun samples and the extraction blank controls.

Plague detection

A concordant plague classification occurred at the species level only in Sk.36 S2, at the species complex level in Sk.36 S1 and Sk 36.S2, and the genus *Yersinia* was additionally present in Sk.45 S2 (Table S4). In the negative controls, no species within the genus *Yersinia* could be identified, nor any taxa within the family *Yersiniaceae* as a whole. It is therefore unlikely that detection of *Y. pestis* can be attributed to reagent contamination. But while detection of the species complex in which *Y. pestis* was found is encouraging, it is ultimately not diagnostic as this complex also contains environmental species, and there is insufficient data present in the shotgun samples to improve upon confidence levels.

Yersinia pestis *genome enrichment*

Whole genome enrichment for *Y. pestis* had a pronounced effect on the taxonomic composition of Sk.36, transforming the *Y. pseudotuberculosis* complex abundance from nearly indistinguishable from sequencing noise to the second-most abundant taxa (Figure 4). Sensitivity at the species level also improved, with identifiable *Y. pestis* becoming the third most abundant taxa. It is curious to note that the abundance of *Homo sapiens* also rose, in both Sk.36 and the enriched extraction blank. One explanation may be sequence similarity between the baits and the human genome, possibly deriving from reference genome contamination. Given the high abundance of *Homo sapiens,* enriched molecules for Sk.36 were also aligned to the human mitochondrial reference sequence rCRS (RefSeq Accession: NC_012920.1) to check for modern contamination. A comparison of the fragment length distribution and terminal cytosine deamination from the reference genome mapping suggests both human and plague molecules are ancient in origin, although there is substantial noise in the plots given the small amount of data used as input (Figure S4). In contrast, the enriched extraction blank contains negligible levels of C->T transitions on the terminal ends of human molecules suggesting little to no evidence of dominant taxa moving from ancient samples to negative controls and vice versa (Table S3). Furthermore, no taxa within the family *Yersiniaceae* were identified in the enrichment blank, affirming the provenience of ancient *Y. pestis* in Sk.36.

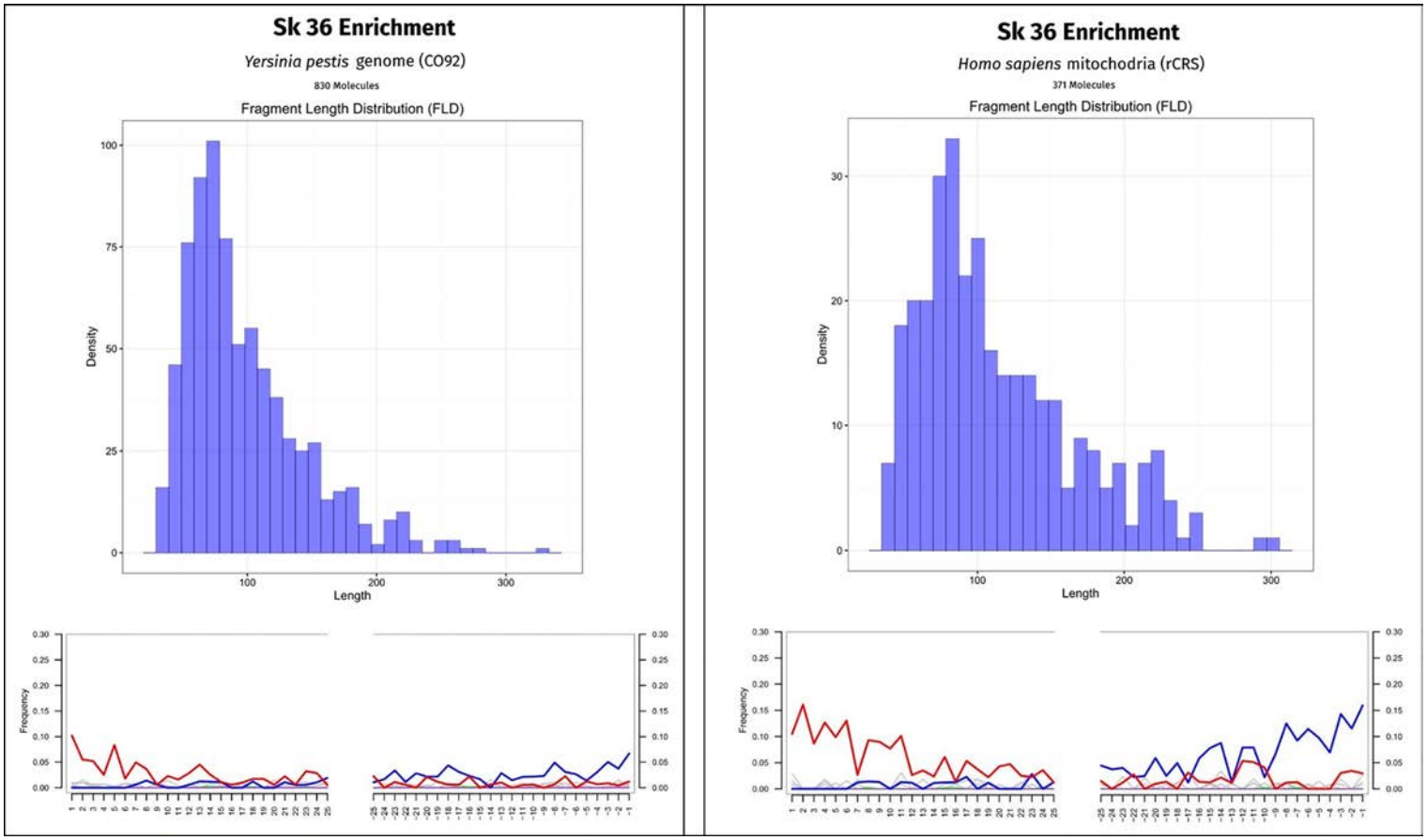

Figure S4. Fragment length distribution and transitions rates for Yersinia pestis *and* Homo sapiens *(red: C->T, blue: G->A).*

The marker gene analysis output of MetaPhlAn2 indicated that *Y. pestis* was the only taxa that could be detected based on alignments to 17 marker gene regions. However, no gene was covered by more than one molecule, as expected for such low sequencing depth and abundance, and thus this finding is conclusive based on the exclusivity of the species detection, rather than the quantitative count of marker genes.

References

ALTSCHUL, S.F., W. GISH, W. MILLER, E.W. MYERS & D.J. LIPMAN. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2

DABNEY, J., M. KNAPP, I. GLOCKE, M.-T. GANSAUGE, A. WEIHMANN, B. NICKEL, C. VALDIOSERA, N. GARCÍA, S. PÄÄBO & J.-L. ARSUAGA. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the USA* 110: 15758–63. https://doi.org/10.1073/pnas.1314445110

JÓNSSON, H., A. GINOLHAC, M. SCHUBERT, P. JOHNSON & L. ORLANDO. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29:1682–84. https://doi.org/10.1093/bioinformatics/btt193

KIRCHER, M., S. SAWYER & M. MEYER. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research.* 40: e3. https://doi.org/10.1093/nar/gkr771

ONDOV, B.D., N.H. BERGMAN & A.M. PHILLIPPY. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12: 385. https://doi.org/10.1186/1471-2105-12-385

RENAUD, G., U. STENZEL & J. KELSO. 2014. leeHom: adaptor trimming and merging for illumina sequencing reads. *Nucleic Acids Research* 42: e141. https://doi.org/10.1093/nar/gku699

SCHWARZ, C., R. DEBRUYNE, M. KUCH, E. MCNALLY, H. SCHWARCZ, A.D. AUBREY, J. BADA & H. POINAR. 2009. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Residues.* 37: 3215–29. https://doi.org/10.1093/nar/gkp159

SUZUKI, R. & H. SHIMODAIRA. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22: 1540–42.

https://doi.org/10.1093/bioinformatics/btl117

TRUONG, D.T., E.A. FRANZOSA, T.L. TICKLE, M. SCHOLZ, E.P. WEINGART, A. TETT, C. HUTTENHOWER & N. SEGATA. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods* 12: 902–903. https://doi.org/10.1038/nmeth.3589

WAGNER, D.M., *et al*. 2014. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet Infectious Diseases* 14: 319–26. https://doi.org/10.1016/S1473-3099(13)70323-2

WEYRICH, L., A.G. FARRER, R. EISENHOFER, L.A. ARRIOLA, J. YOUNG, A.S. CAITLIN, M. HANDSLEY-DAVIS, C. ADLER, J. BREEN & A. COOPER. 2019. Laboratory contamination over time during low-biomass sample analysis. *Molecular Ecology Resources* 19: 982–96. https://doi.org/10.1101/460212

WICKHAM, H. 2009. *ggplot2: elegant graphics for data analysis*. New York: Springer. https://doi.org/10.1007/978-0-387-98141-3

WOOD, D.E. & S.L. SALZBERG. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* 15: R46. https://doi.org/10.1186/gb-2014-15-3-r46

Table S1. Sample overview and *pla* PCR screening results; highlighted rows refer to the two samples positive for *Yersinia pestis.*

Primary ID	Sample type	Secondary ID	Descriptions	*pla* PCR round 1	*pla* PCR round 2
Thor13, Sk.1	Insufficient root	HP2645	Lower LM1 (19)	--	--
Thor13, Sk.2	Molar root	HP2646	Upper RM1 (3)	0/2	NA
Thor13, Sk.3	Molar root	HP2647	Upper RM1 (3)	0/2	NA
Thor13, Sk.4	Molar root	HP2648	Upper LM2 (15)	0/2	NA
Thor13, Sk.6	Molar root	HP2649	Lower LM2 (18)	0/2	NA
Thor13, Sk.9	Molar root	HP2650	Lower LM3 (17)	0/2	NA
Thor13, Sk.10	Molar root	HP2651	Lower RM3 (32)	0/2	NA
Thor13, Sk.11	Insufficient root	HP2652	Lower RM2 (31)	--	--
Thor13, Sk.12	Insufficient root	HP2653	Lower LM1 (19)	--	--
Thor13, Sk.14	Molar root	HP2654	Upper LM3 (16)	0/2	NA
Thor13, Sk.17	Insufficient root	HP2655	Upper LM3 (16)	--	--
Thor13, Sk.18	Insufficient root	HP2656	Lower RM1 (30)	--	--
Thor13, Sk.19	Molar root	HP2657	Upper RM3 (1)	0/2	NA
HP2658	Insufficient root	HP2658	Upper LM1 (14)	--	--
Thor13, Sk.25	Molar root	HP2659	Upper LM3 (16)	0/2	NA
Thor13, Sk.28	Molar root	HP2660	Lower RM3 (32)	0/2	NA
Thor13, Sk. 35	Molar root	HP2661	Lower Ldm2 (61)	0/2	NA
Thor13, Sk.36	Molar root	HP2662	Lower Ldm1 (62)	2/2	6/6
Thor13, Sk.38	Molar root	HP2663	Lower Rdm1 (69)	0/2	NA
Thor13, Sk.42	Insufficient root	HP2664	Lower RM3 (32)	--	--
Thor13, Sk.43	Molar root	HP2665	Lower RM1 (30)	0/2	NA
Thor13, Sk.45	Molar root	HP2666	Lower RM1 (30)	1/2*	0/4
Thor13, Sk.46	Insufficient root	HP2667	Lower LM2 (18)	--	--
Thor13, Sk.47	Insufficient root	HP2668	Lower LM2 (18)	--	--
Thor13, Sk.49	Molar root	HP2669	Lower LM3 (17)	0/2	NA
EBC S1	Extraction blank control 1	NA	DNA extraction	0/2	NA
EBC S2	Extraction blank control 2	NA	DNA extraction	0/2	NA
EBC E	Extraction blank control	NA	DNA extraction	0/2	NA

	(enriched)				
LBC S1	Library blank control 1	NA	Library prep	NA	NA
LBC S2	Library blank control 2	NA	Library prep	NA	NA

* 5’ C -> T / 3’ G -> A transitions were calculated from the 5 bp termini of the blast alignments.

Table S2. Sequencing and general metagenomic classification statistics.

ID	Sequence round	Sequenced molecules	Pre-processed molecules	Metagenomic subsample	Species concordant molecules	Number of species
Sk.36	S1	2 060 978	1 746 908	250 000	541	269
Sk.36	S2	3 383 126	2 806 528	250 000	433	227
Sk.45	S1	1546	1060	1060	25	12
Sk.45	S2	5 487 465	3 938 696	250 000	1098	331
Sk.36	E	174 591	36 519	36 519	6589	33
EBC	S1	6130	616	616	68	14
EBC	S2	38 867	6237	6237	409	86
EBC	E	4398	675	675	300	21
LBC	S1	0	0	0	0	0
LBC	S2	21 712	1218	1218	61	22

Table S3. *Homo sapiens* metagenomic classification statistics; highlighted rows refer to the samples positive for *Yersinia pestis.*

ID	Sequence Round	% *Homo sapiens*	% terminal C->T	% Non-terminal C->T	Damage ratio
Sk.36	S1	0.01%	2.17	1.44	1.51
Sk.36	S2	0.01%	2.63	2.23	1.18
Sk.45	S1	0.38%	0	0.90	0.00
Sk.45	S2	0.15%	4.19	0.90	4.66
Sk.36	E	8.60%	3.57	2.53	1.41
EBC	S1	0.49%	0	0	0.00
EBC	S2	1.04%	0	0.43	0.00
EBC	E	30.52%	0	0.15	0.00
LBC	S1	0	0	0	0.00

LBC	S2	2.30%	6.67	1.78	3.75

Table S4. *Yersinia* metagenomic classification statistics.

ID	Sequence round	Sequenced molecules	Pre-Processed molecules	*Yersinia*	*Yersinia pseudotuberculosis complex*	*Yersinia pestis*
Sk.36	S1	2 060 978	1 746 908	6	2	0
Sk.36	S2	3 383 126	2 806 528	14	9	1
Sk.45	S1	1546	1060	0	0	0
Sk.45	S2	5 487 465	3 938 696	26	0	0
Sk.36	E	174 591	36 519	4034	3181	599
EBC	S1	6130	616	0	0	0
EBC	S2	38 867	6237	0	0	0
EBC	E	4398	675	0	0	0
LBC	S1	0	0	0	0	0
LBC	S2	21 712	1218	0	0	0

Table S5. The top 10 most abundant species in the ancient sample shotgun cluster.

Sk36S1		Sk36S2		Sk45S2	
Species	**Ecology**	**Species**	**Ecology**	**Species**	**Ecology**
Luteitalea pratensis	Soil	*Homo sapiens*	Human	*Homo sapiens*	Human
Homo sapiens	Human	*Luteitalea pratensis*	Soil	*Steroidobacter denitrificans*	Water
Rhodoplanes sp. Z2-YC6860	Water	*Rhodoplanes sp. Z2-YC6860*	Water	*Bradyrhizobium icense*	Plant-associated
Gemmatirosa kalamazoonesis	Soil	*Steroidobacter denitrificans*	Water	*Luteitalea pratensis*	Soil
Steroidobacter denitrificans	Water	*Gemmatirosa kalamazoonesis*	Soil	*Betaproteobacteria bacterium GR16-43*	Water
Lysobacter enzymogenes	Soil/Water	*Streptomyces glaucescens*	Soil	*Nitrospira moscoviensis*	Water
Nonomuraea sp.	Soil	*Nitrospira moscoviensis*	Water	*Ramlibacter tataouinensis*	Soil
Nitrospira defluvii	Water	*Nitrospira japonica*	Water	*Gemmatirosa kalamazoonesis*	Soil
Bradyrhizobium icense	Plant-associated	*Nitrospira defluvii*	Water	*Rhodoplanes sp. Z2-YC6860*	Water
Xanthomonas translucens	Plant-associated	*Streptomyces lunaelactis*	Soil/Water	*Bradyrhizobium erythrophlei*	Plant-associated

Table S6. The top 10 most abundant species in the extraction blank control and ethanol-inhibited cluster.

Sk45S1		EBCS1		EBCS2	
Species	**Ecology**	**Species**	**Ecology**	**Species**	**Ecology**
Desulfobulbus sp. ORNL	Human oral	*Bradyrhizobium sp. SK17*	Soil	*Homo sapiens*	Human
Homo sapiens	Human	*Homo sapiens*	Human	*Bradyrhizobium sp. SK17*	Soil
Eubacterium minutum	Human oral	*Alcaligenes faecalis*	Soil/faeces	*Sphingopyxis terrae*	Water
Actinomyces sp. oral taxon 414	Human oral	*Variovorax sp. HW608*	Soil	*Acidovorax sp. KKS102*	Water
Actinomyces hongkongensis	Human blood	*Treponema denticola*	Human oral	*Cupriavidus metallidurans*	Water
Tannerella forsythia	Human oral	*Stenotrophomonas maltophilia*	Water	*Bosea sp. AS-1*	Soil/water
Ramlibacter tataouinensis	Soil	*Rhodoplanes sp. Z2-YC6860*	Soil	*Acidovorax sp. RAC01*	Water
Olsenella sp. oral taxon 807	Human oral	*Pseudomonas pseudoalcaligenes*	Soil/water	*Ralstonia pickettii*	Soil/Water
Nonomuraea sp.	Soil	*Comamonas testosteroni*	Soil/human-Associated	*Acidovorax carolinensis*	Soil
Lachnospiraceae bacterium oral taxon 500	Human oral	*Bradyrhizobium sp. BTAi1*	Plant-associated	*Pseudomonas aeruginosa*	Soil/water/human-Associated

Table S7. The top ten most abundant species in the enrichment and library blank cluster.

Sk36E		EBCE		LBCS2	
Species	**Ecology**	**Species**	**Ecology**	**Species**	**Ecology**
Homo sapiens	Human	*Homo sapiens*	Human	*Homo sapiens*	Human
Yersinia pseudotuberculosis complex	Soil/water				
Human-associated	*Alcaligenes faecalis*	Soil/faeces	*Bradyrhizobium* sp. SK17	Soil	
Yersinia pestis	Soil/Water/				
Human-associated	*Klebsiella aerogenes*	Soil/water	*Cutibacterium acnes*	Human-associated	
Plant-associated					
Steroidobacter denitrificans	Water	*Micrococcus luteus*	Soil/water/human skin	*Diaphorobacter polyhydroxybutyrativorans*	Water
Wenzhouxiangella marina	Water	*Sphingopyxis terrae*	Water	*Acidovorax carolinensis*	Soil
Nitrospira moscoviensis	Water	*Bradyrhizobium* sp. SK17	Soil	*Xanthomonas campestris*	Plant-associated
Nitrospira japonica	Water	*Acidovorax* sp. KKS102	Water	*Tannerella forsythia*	Human oral
Sulfuricaulis limicola	Water	*Cupriavidus metallidurans*	Water	*Streptomyces* sp. 3214.6	Soil/water
Gemmatirosa kalamazoonesis	Soil	*Acinetobacter radioresistens*	Soil/water	*Streptococcus mitis*	Human oral