Article

Assessing Metagenomic Signals Recovered from Lyuba, a 42,000-Year-Old Permafrost-Preserved Woolly Mammoth Calf

Giada Ferrari [1,2,*] ![ID], **Heidi E. L. Lischer** [3,4], **Judith Neukamm** [1,5] ![ID], **Enrique Rayo** [1], **Nicole Borel** [6], **Andreas Pospischil** [6], **Frank Rühli** [1], **Abigail S. Bouwman** [1] **and Michael G. Campana** [1,7,*] ![ID]

[1]   Institute of Evolutionary Medicine, University of Zurich, 8057 Zurich, Switzerland; judith.neukamm@iem.uzh.ch (J.N.); enrique.rayo@iem.uzh.ch (E.R.); frank.ruehli@iem.uzh.ch (F.R.); abigail.bouwman@iem.uzh.ch (A.S.B.)
[2]   Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, 0316 Oslo, Norway
[3]   Institute of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland; heidi.lischer@ieu.uzh.ch
[4]   Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland
[5]   Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany
[6]   Institute of Veterinary Pathology, University of Zurich, 8057 Zurich, Switzerland; n.borel@access.uzh.ch (N.B.); apos@vetpath.uzh.ch (A.P.)
[7]   Center for Conservation Genomics, Smithsonian Conservation Biology Institute, Washington, DC 20008, USA
*   Correspondence: giada.ferrari@ibv.uio.no (G.F.); campanam@si.edu (M.G.C.); Tel.: +47-22-85-50-65 (G.F.); +1-202-633-4183 (M.G.C.)

![check for updates]

Abstract: The reconstruction of ancient metagenomes from archaeological material, and their implication in human health and evolution, is one of the most recent advances in paleomicrobiological studies. However, as for all ancient DNA (aDNA) studies, environmental and laboratory contamination need to be specifically addressed. Here we attempted to reconstruct the tissue-specific metagenomes of a 42,000-year-old, permafrost-preserved woolly mammoth calf through shotgun high-throughput sequencing. We analyzed the taxonomic composition of all tissue samples together with environmental and non-template experimental controls and compared them to metagenomes obtained from permafrost and elephant fecal samples. Preliminary results suggested the presence of tissue-specific metagenomic signals. We identified bacterial species that were present in only one experimental sample, absent from controls, and consistent with the nature of the samples. However, we failed to further authenticate any of these signals and conclude that, even when experimental samples are distinct from environmental and laboratory controls, this does not necessarily indicate endogenous presence of ancient host-associated microbiomic signals.

Keywords: ancient DNA; *Mammuthus primigenius*; microbiome; environmental DNA; DNA contamination

1. Introduction

The field of paleomicrobiology has experienced significant advances thanks to the development of high-throughput DNA sequencing techniques: From the first attempts to extract ancient DNA (aDNA) from mummified tissues [1], to validating the presence of pathogens [2–4], to reconstructing ancient bacterial genomes [5,6], to recovering entire ancient microbial communities [7,8]. Given the importance

of commensal microbes in human health [9–11], increasing numbers of microbiome studies have been conducted in recent years. Consequently, much attention has also been given to the taxonomic reconstruction of ancient microbiomes, with a particular focus on human fecal and oral microbiomes using coprolites [12] and dental calculus [7,8,13]. Such approaches offer an ideal opportunity to observe the evolution of these site-specific microbial communities over time.

Since the majority of ancient microbiome studies rely on archaeological material, particular attention needs to be given to the issue of environmental exposure to closely related soil taxa. The presence of DNA from unsequenced or poorly characterized environmental microorganisms potentially, related to pathogens can confound paleomicrobiological studies [14,15]. Such issues are even more prominent when attempting to reconstruct entire ancient microbial communities and authentication procedures need therefore to address environmental signals [16,17]. Additionally, the same issues can be caused by contaminants in laboratory reagents [18–20].

Here, we illustrate the issue of environmental and laboratory contamination by attempting to reconstruct tissue-specific ancient metagenomes from a permafrost-preserved woolly mammoth (*Mammuthus primigenius*, Blumenbach 1799) calf. Asian elephants (*Elephas maximus*), the closest living relatives of woolly mammoths [21], are known to be affected by a number of microbial diseases, such as elephant endotheliotropic herpesviruses [22] or tuberculosis [23]. We therefore set out to investigate whether this was the case for woolly mammoths too, and chose a shotgun metagenomics approach, in order to capture the entire microbial diversity.

The calf, originally named "Lyuba", was discovered by Nenets reindeer herders in 2007, exposed on a point bar along the Yuribey river on the Yamal peninsula, northwestern Siberia (68°38′ N, 71°40′ E) [24,25]. Lyuba (Figure 1) is the best preserved mummified woolly mammoth ever found, presenting only slight skin lesions [24]. Radioisotopic analyses of bone collagen indicate that it lived approximately 41,800 ^{14}C years BP [25], and anatomical investigations revealed that the calf was female and in a good nutritional state [24,25]. The age at death was estimated at 30–35 days, based on a dentin neonatal line and postnatal dentin increments on its deciduous teeth [26]. Vivianite (hydrated iron phosphate), was found in Lyuba's trachea and bronchi, suggesting the cause of death to be suffocation after inhalation of mud [25]. Due to Lyuba's accidental cause of death, good health, and excellent preservation, several studies have been performed on its remains, since findings can potentially be extended to juvenile woolly mammoths in general.

Figure 1. Lateral right view of Lyuba (photo credit: Daniel C. Fisher. Museum of Paleontology, University of Michigan, Ann Arbor, MI, USA). From Reference [27], reprinted by permission of John Wiley & Sons, Inc. (Hoboken, NJ, USA).

Fisher and colleagues [25] argue that the point bar on which Lyuba was found was not likely to represent the site of death or preservation. They conclude that an ice-out flooding during the previous summer is the most likely natural process to have moved the calf to its discovery location. Therefore, Lyuba must have endured at least one cycle of thawing and freezing before discovery, a finding that has also been corroborated by histological analysis [27].

Dissections were performed on Lyuba in 2008 and 2009 [25], and several samples, including intestinal, peritoneal, muscle, and fat tissues, and vivianite, were obtained for further analysis. Here, we attempted to reconstruct these samples' metagenomes through shotgun high-throughput DNA sequencing. Despite displaying a large component derived from environmental and laboratory contamination, the tissue-specific metagenomes were taxonomically different from non-template and environmental controls. Furthermore, some bacterial species were only associated with one specific tissue, consistent with a potential bacterial infection or a tissue-specific microbiomic signal. However, we failed to further authenticate any of these signals and conclude that, in this case, all detected signals were the result of laboratory or environmental contamination. Caution and strict validation procedures are therefore recommended even under ideal morphological preservation circumstances and when comparison with laboratory and environmental control samples suggest endogenous presence of microbial species.

2. Materials and Methods

2.1. Samples

The samples used in this study (donated by Daniel C. Fisher, Museum of Paleontology, University of Michigan, 1109 Geddes Ave., Ann Arbor, MI 48109-1079, USA) were obtained during endoscopical examination, and partial dissections performed in 2008 and 2009, and described in References [25,27]. The authors were not present during these procedures, but received the samples at a later stage. This study was conducted with written permission of the sample donors and of the International Mammoth Committee. Lyuba was kept frozen during the time intervals between discovery and the two dissections. The samples used in this study were previously labeled as follows: Abdominal oblique muscle, abdominal subcutaneous fat and muscle, intestinal tissue (two samples), caecum, omentum, peritoneum, cheek fat, and vivianite.

2.2. DNA Extraction

The samples' surfaces were cleaned with a 1% NaOCl solution and UV-irradiated for 10 min to reduce contamination. One hundred milligrams of homogenized tissue were digested at 55 °C for 18 h, followed by three days rotating at room temperature, in 1 mL of extraction buffer (50 mM Tris-HCl, 100 mM NaCl, 25 mM EDTA, 2% SDS and 0.5 mg/mL proteinase K). Following centrifugation, the supernatant was extracted twice with a 25:24:1 phenol, chloroform, and isoamyl alcohol mixture, followed by a final chloroform step. DNA was isolated using QIAquick spin columns (QIAGEN, Hombrechtikon, Switzerland), with two elutions in 30 μL EB buffer and reduced centrifugation speed (6–10 krpm) to prevent loss of short DNA fragments. Extractions were performed in a dedicated aDNA clean laboratory at the University of Zurich [28], which has an independent HEPA (High Efficiency Particulate Air) air filtration system, is physically separated from all laboratories in which PCR is performed, and follows established anti-contamination protocols [29,30], including, unidirectional work flows, overhead UV lights, regular sterilization of all work surfaces, the use of full body suits, masks, and gloves by all researchers, and the parallel processing of non-template controls.

2.3. Library Preparation and Sequencing

Ten microliters of extract or extraction blank were converted into double-indexed Illumina sequencing libraries following a protocol optimized for aDNA [31,32]. While all samples had unique index pairs, some shared a single index. Sample-specific indexes were added to each library via

10 cycles of PCR amplification [32]. Blunt-end repair, adapter ligation and set up of indexing PCRs were performed in an aDNA clean laboratory and non-template library blanks were generated in parallel. In order to increase library concentrations to 10^{13} copies/µL, a re-amplification step was performed on all indexed libraries in 100 µL reactions containing 1 unit AccuPrime™ *Pfx* DNA polymerase (Thermo Fisher Scientific, Reinach, Switzerland), 1× AccuPrime™ *Pfx* reaction mix, 0.3 µM primers IS5 and IS6 [31] and 0.5–4 µL library template with the following thermal profile: Initial denaturation at 95 °C for 2 min, 9–18 cycles of denaturation at 95 °C for 15 s, annealing at 60 °C for 30 s and elongation at 68 °C for 1 min, followed by a final elongation at 65 °C for 5 min. Libraries were purified with MinElute spin columns (QIAGEN, Hombrechtikon, Switzerland) following the manufacturer's instructions. Quantitative PCR and analysis on an Agilent 2100 Bioanalyzer DNA 1000 chip were used to assess library qualities. Libraries were pooled equally and sequenced on a MiSeq (Illumina, San Diego, CA, USA) with paired-end 150 bp reads, as well as on 1.5 lanes of a HiSeq2500 (Illumina) with paired-end 125 bp reads and v4 chemistry by the Functional Genomics Center Zurich (Zurich, Switzerland). While all samples were sequenced on the same MiSeq lane, individuals which shared an index were separated across the two HiSeq lanes to limit index switching [32].

2.4. Sequence Quality Control and Filtering

All Illumina paired-end reads (MiSeq and HiSeq) were processed using the same pipelines. Raw reads were demultiplexed using DeML [33] with default settings. Due to low-quality p5 sequences, the HiSeq data were demultiplexed using only the p7 adapter sequence. Trimmomatic 0.33 [34] was used to remove adapter sequences and artifacts with the following parameters: Maximum seed mismatches = 2, palindrome clip threshold = 30, simple clip threshold = 7, minimum adapter length = 5, and retaining reverse reads. Leading and trailing bases below Phred quality score 3 were removed, reads were scanned using 4 bp sliding windows and trimmed when average Phred quality score fell below 15. Trimmed reads shorter than 25 bp were discarded. Surviving paired reads were merged with SeqPrep [35] as described in Reference [36] to increase sequence quality. Merged reads shorter than 25 bp and unmerged reads were discarded. Finally reads qualities were assessed using FastQC 0.10.1 [37].

2.5. Mapping to Reference Genomes and Deamination Pattern Analysis

Reads were mapped to the African elephant (*Loxodonta africana*) nuclear genome (NW_003573420.1), the woolly mammoth (*M. primigenius*) mitochondrial genome (EU155210.1), and the human reference genome (GRCh38.p2) using the Burrows-Wheeler Aligner (BWA) 0.7.12 aln algorithm [38], with disabled seeding, increased gap open (-o 2), and reduced edit distance (-i 0) as recommended in Reference [39]. Duplicates were removed with Picard 2.1.0 [40] with validation stringency set to lenient, and filtering for a minimal mapping quality of 25 was performed with SAMtools 1.3.1 [41]. DNA deamination patterns of mapped reads were analyzed with mapDamage2 2.0.6 [42].

2.6. Mitogenome Analysis

MiSeq and HiSeq reads, that mapped to the woolly mammoth mitochondrial genome for all samples, excluding vivianite, were merged into a single BAM file, duplicates were marked as described above and a consensus sequence was built using SAMtools 1.3.1 [41]. LASTZ [43] was used to compare the obtained consensus sequence with the Lyuba mitochondrial genome published by Enk and colleagues [44] (KX027526.1).

2.7. Shotgun Metagenomic Analysis

From this step onwards, published Illumina shotgun metagenomes from two Asian elephant (*E. maximus*) fecal samples (SRP040073) [45] and two Russian permafrost samples (SRP049520) [46] obtained from the Sequence Read Archive were included. The elephant fecal samples were obtained from three-week-old and six-year-old Asian elephants (*E. maximus*). No other comparable elephant metagenomic datasets are currently available. The permafrost samples originated from

lake sediment (alluvium) from the Panteleikha River floodplain and from the late Pleistocene Ice Complex on the Omolon River, both located in northeast Siberia. These reads were processed using the same parameters as for the Lyuba samples. Quality-controlled, merged reads from the libraries generated here and the reference data were deduplicated by clustering and removing all exact matches using CD-HIT-EST 4.6 [47]. Deduplicated reads were compared to the National Center for Biotechnology Information (NCBI) nucleotide collection (downloaded January 2018) using the megablast algorithm (BLAST 2.6.0+ [48]) with default parameters. To verify megablast species identifications, we additionally analyzed the Lyuba, vivianite, and laboratory controls with MALT [49] using a curated reference database consisting of 5242 complete bacterial genomes available in NCBI RefSeq (December 2015). The BlastN algorithm with SemiGlobal alignment was used. Megablast and MALT results were visualized with MEGAN6 (builds 6.11.6 or 6.10.6, respectively) [50]. The megablast results were analyzed using MEGAN default Lowest Common Ancestor (LCA) parameters, while the MALT results used the following options: –topPercent 1.0, –minSupportPercent 0, and –minSupport 5. Sample comparisons used normalized counts to control for variation in sequencing depth.

2.8. 16S Metagenomic Analysis

16S metagenomic analysis was performed using the QIIME2 (version 2018.4) [51] package and its associated plugins. All analyses were performed using default settings unless otherwise specified. Deduplicated reads from the novel and published shotgun libraries were closed-reference clustered at 99% sequence identity to the SILVA database release 132 [52] using VSEARCH [53]. Retained sequences were aligned and a phylogenetic tree was constructed using MAFFT [54] and FastTree [55]. The library blank was excluded from diversity metric calculations because only nine sequences were retained after clustering. Phylogenetic and non-phylogenetic diversity metrics were calculated with rarefaction to 149 and 945 sequences, corresponding to the next two smallest sample sizes (extraction blank and intestinal tissue 2, respectively) from our novel Lyuba dataset. Principal Coordinate Analyses of Bray-Curtis, Jaccard, and Unweighted UniFrac distances were visualized using EMPeror [56,57].

Clustered sequences were classified according to the SILVA consensus taxonomy (all levels) using VSEARCH [53] (99% nucleotide identity). Clustered sequences were exported from QIIME 2 and analyzed using SourceTracker 2 [58]. All mammoth tissue samples were classified as 'sinks', while the permafrost, blanks, feces and vivianite samples were treated as 'sources'. Due to the very small sample sizes, we did not rarefy the sources and rarefied the sinks to 100. All other parameters used their default values.

2.9. MetaPhlAn2 Analysis

We also analyzed the deduplicated metagenomes using MetaPhlAn2 (version 2.7.7) under default settings [59]. We generated heat maps clustering the individuals and observed species-level taxa using Euclidean distances.

2.10. Authentication of Bacterial Reads

In order to test the authenticity of overrepresented bacterial species in the omentum and cheek fat samples, reads were mapped against the reference genomes of several *Yersinia* species for the omentum sample (*Y. aldovae*, NZ_CP009781.1; *Y. aleksiciae*, NZ_CP011975.1; *Y. enterocolitica*, NC_008800.1; *Y. entomophaga*, NZ_CP010029.1; *Y. frederiksenii*, NZ_CP009364.1; *Y. intermedia*, NZ_CP009801.1; *Y. kristensenii*, NZ_CP008955.1; *Y. pestis*, NC_003143.1; *Y. pseudotuberculosis*, NZ_CP008943.1; *Y. rohdei*, NZ_CP009787.1; *Y. ruckeri*, NZ_CP011078.1; and *Y. similis* NZ_CP007230.1), as well as to the *Alcaligenes faecalis* reference genome (NZ_CP013119.1) and the *Carnobacterium* sp. CP1 genome (NZ_CP010796.1) for the cheek fat sample, using the same parameters as described above. DNA deamination patterns of mapped reads were analyzed with mapDamage2 2.0.6 [42].

2.11. Data Availability

Raw sequencing data have been deposited in the NCBI Sequence Read Archive (accession number SRP113695).

3. Results

3.1. Data Quality Control and Authentication

We obtained 17–47 million reads per sample and, with the exception of the vivianite sample (39% reads surviving), 81–93% of the raw reads survived quality filtering and merging (Table 1). Average read lengths ranged from 34 to 56 bp, with only few reads longer than 100 bp (Figure 2), consistent with fragmentation typical for aDNA [60].

(**a**)

(**b**)
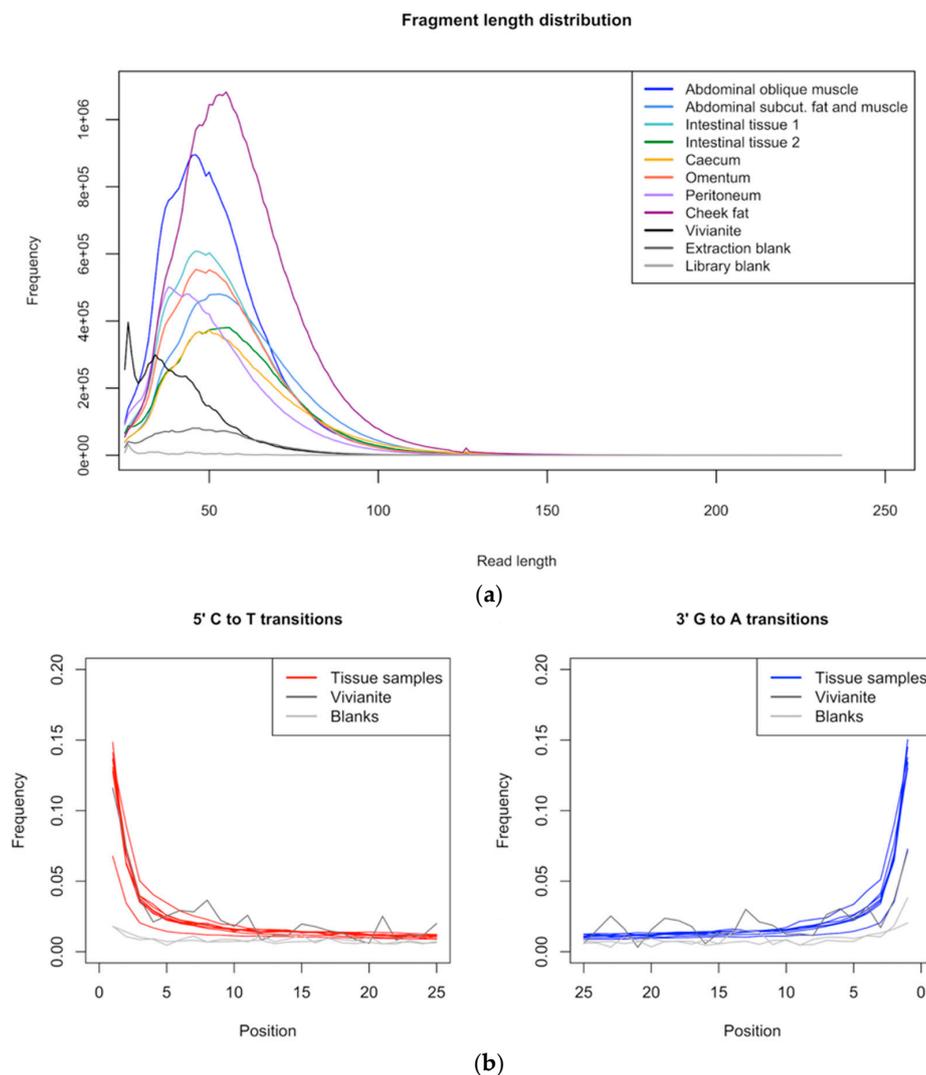
Figure 2. Ancient DNA (aDNA) fragmentation and misincorporation patterns. (**a**) Read length distribution of quality-filtered reads is shown for all samples and controls. Average read length ranges from 34 to 56 bp. (**b**) Quality-filtered reads were mapped to the African elephant nuclear genome with a minimum mapping quality of 25 and nucleotide misincorporation rates were calculated using mapDamage. Increased cytosine deamination rates at 5′-overhangs are visible (C to T and G to A transitions), consistent with aDNA.

Quality-filtered, merged reads were mapped to the African elephant (*L. africana*) nuclear genome and the human reference genome, as well as the woolly mammoth (*M. primigenius*) mitogenome. Tissue samples showed a highly variable number of reads mapping to the elephant and mammoth references (2.1–43% and 0.1–0.8%, respectively), indicating different levels of endogenous mammoth DNA preservation depending on the tissue of origin (Table 1). The abdominal oblique muscle and cheek fat samples yielded the highest proportions of reads mapping to the elephant nuclear genome (42–43% of total unique reads), whereas for the vivianite sample less than 2500 reads mapped to the elephant nuclear genome (0.28% of total unique reads). Since the vivianite sample did not differ significantly from negative controls in terms of raw data quality and endogenous DNA content, we treated it as an environmental control sample. mapDamage analysis of reads mapped to the African elephant nuclear genome (Figure 2) and woolly mammoth mitochondrial genome (Figure S1a) yielded DNA damage patterns consistent with authentic aDNA, showing increased cytosine deamination rates at 5′-overhangs for all tissue samples [42].

Table 1. Quality filtering and mapping statistics for reads obtained from shotgun sequencing.

	Raw Read Pairs	Merged Reads (% of Total)	Unique Reads (Clonality)	Unique Mapped Reads (% of Unique Reads)		
				Loxodonta africana	*Mammuthus primigenius*	*Homo sapiens*
				Nuclear Genome	Mitogenome	
Abdominal oblique muscle	35,511,605	31,176,532 −87.80%	102,34,165 (3.05×)	4,371,482 −42.71%	55532 −0.54%	78047 −0.76%
Abd. subcut. fat and muscle	22,977,668	21,176,646 −92.20%	6,834,930 (3.10×)	143221 −2.10%	16216 −0.24%	20252 −0.29%
Intestinal tissue 1	25,053,246	23,305,474 −93.00%	7,496,017 (3.11×)	605598 −8.08%	32548 −0.43%	16223 −0.22%
Intestinal tissue 2	21,130,716	17,066,995 −80.80%	2,829,012 (6.03×)	270976 −9.58%	12162 −0.43%	6882 −0.24%
Caecum	17,386,802	16,168,853 −93.00%	5,819,269 (2.78×)	634528 −10.90%	46103 −0.79%	12498 −0.21%
Omentum	22,924,111	21,390,754 −93.30%	6,909,388 (3.10×)	186132 −2.69%	9453 −0.14%	7581 −0.11%
Peritoneum	18,925,818	17,416,113 −92.00%	5,994,320 (2.91×)	405067 −6.76%	6093 −0.10%	14650 −0.24%
Cheek fat	47,716,128	42,628,440 −89.30%	11,029,081 (3.87×)	1947669 −42.08%	8582 −0.19%	37331 −0.81%
Vivianite	21,888,799	8,481,750 −38.80%	824,368 (10.29×)	2334 −0.28%	194 −0.02%	6298 −0.76%
Extraction blank	7,036,101	3,224,382 −45.80%	711,189 (4.53×)	22774 −3.20%	227 −0.03%	5038 −0.71%
Library blank	4,820,063	285,265 −5.90%	33387 (8.54×)	8417 −25.21%	151 −0.45%	4479 −13.42%

We also observed a low degree of sample contamination with human DNA (0.11–0.81%, Table 1). DNA damage patterns showed a slight increase in cytosine deamination rates at 5′-overhangs (Figure S1b), suggesting that contamination with human DNA happened prior to this study and is consistent with the extensive handling of Lyuba at discovery and recovery. We detected small numbers of unique reads that matched elephant, mammoth and humans in the library and extraction blanks. Since some samples shared indexes with the blanks, these likely derive from sample mis-assignment due to index switching [32].

In order to further validate the authenticity of the sequencing data, we pooled all reads mapping to the woolly mammoth mitogenome together, remapped them to the reference, and built a consensus sequence for the mitogenome of Lyuba. We obtained a sequence with an average per-base coverage of 227×, which was consistent with the previously published Lyuba mitogenome [44].

3.2. Shotgun Metagenomic Analysis

In order to reconstruct the taxonomic composition of Lyuba's tissue-specific metagenomes, we compared quality-filtered, deduplicated reads to the NCBI nucleotide database with megablast and visualized the results with MEGAN6 [50]. Additionally, we included published metagenomes obtained from Asian elephant fecal samples [45] and Russian permafrost sediments [46] for comparison.

For all Lyuba tissue samples, reads could be assigned to the genera *Loxodonta*, *Elephas*, or *Mammuthus* (25,000–600,000 reads per sample), with the abdominal oblique muscle sample yielding the most reads (Figure 3b). Furthermore, as expected for this kind of material, we observed a large environmental DNA component in Lyuba's tissues, with the most commonly occurring bacterial species being known environmental, soil- or water-dwelling, bacteria (Figure 3a,b). For example, roughly 50-80% of the metagenome of all tissue samples, excluding the cheek fat sample, is composed of bacterial species belonging to the genera *Pseudomonas*, *Janthinobacterium*, *Caulobacter*, and *Brevundimonas*. All these taxa are present in the vivianite control, and in at least one of the laboratory non-template controls, suggesting contamination from laboratory reagents or workflows, the environment, or both. Despite this, the remaining fraction of the tissue samples' metagenomes appear to have overall different taxonomic compositions and to differ from the non-template controls as well as from the vivianite control (Figure 3b). This can also be seen in the Principal Coordinate Analysis (PCoA) of taxonomic profiles at the species level (Figure 3c), where the tissue samples cluster separately from the laboratory and vivianite controls and the cheek fat sample is most distant from the other tissue samples. Contrary to expectations, while mammoth samples of intestinal origin cluster together, they are not taxonomically similar to the elephant fecal samples. Similarly, the vivianite sample is taxonomically different from the permafrost samples.

Interestingly, we also observed a few bacterial taxa that were particularly abundant in the metagenome of one of Lyuba's tissue samples and nearly or completely absent from the other tissues and controls. In particular, we detected the presence of *Yersinia* (354,641 summed reads) in the omentum sample and of *Carnobacterium* (307,572 summed reads) and *Alcaligenes* (272,401 summed reads) in the cheek fat sample. While species of the genus *Yersinia* are widely found in the environment, mostly in fresh water and soil, some are important pathogens for humans and other animals (*Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica*) and yet others are capable of opportunistic infections [61]. The majority of reads (330,049) were assigned at the genus level rather than to one or more species in particular. Similarly, *A. faecalis*, to which the majority of the *Alcaligenes* reads (271,878) were assigned to, is a common soil bacterium and human opportunistic pathogen [62]. In contrast, there are no known pathogens in the genus *Carnobacterium* [63]. However, its presence in the cheek fat sample is interesting, since acidification of Lyuba's tissues through lactic-acid-producing bacteria has been suggested as an explanation for its exceptional preservation and lack of scavenging during the time the calf was exposed between the ice-out flooding and its recovery [25]. The majority of reads (201,132) were assigned to *Carnobacterium* sp. CP1.

When ranking bacterial taxa in Lyuba's tissues at the genus and species level and ignoring all taxa that were present in the laboratory or environmental vivianite controls (with a minimum of 50 assigned reads), we observed a much lower diversity in taxonomic composition (Figure S2). Furthermore, all detected species and genera are common soil- or water-dwelling bacteria. The only exception is the marked presence of *Yersinia* in the omentum sample. Because the laboratory controls contained 80–160 reads that were assigned to *Alcaligenes* or *Carnobacterium*, we could not observe these taxa in the cheek fat sample in this analysis. However, given the high number of reads in the tissue sample (272,401 and 307,572 summed reads, respectively), this could also be due to index switching rather than to laboratory contamination.

Finally, we also analyzed all quality-filtered reads obtained from Lyuba's tissue samples and controls with MALT [49] and visualized the results with MEGAN6 [50]. Results were generally consistent with the outcome of the megablast analysis (Figure S3), but we did observe some discrepancies. For example, the MALT analysis did not detect *Janthinobacterium* as a major component

of most of Lyuba's tissues. We also obtained fewer reads assigned to the genus *Yersinia* (95,021 summed reads) in the omentum sample. Furthermore, we did not observe the presence of *A. faecalis* in the cheek fat sample. These discrepancies are due to the fact that the reference database we utilized for the MALT analysis contains fewer *Janthinobacterium* and *Yersinia* spp. genomes and lacks the *A. faecalis* genome.

(**a**)   (**b**)

(**c**)
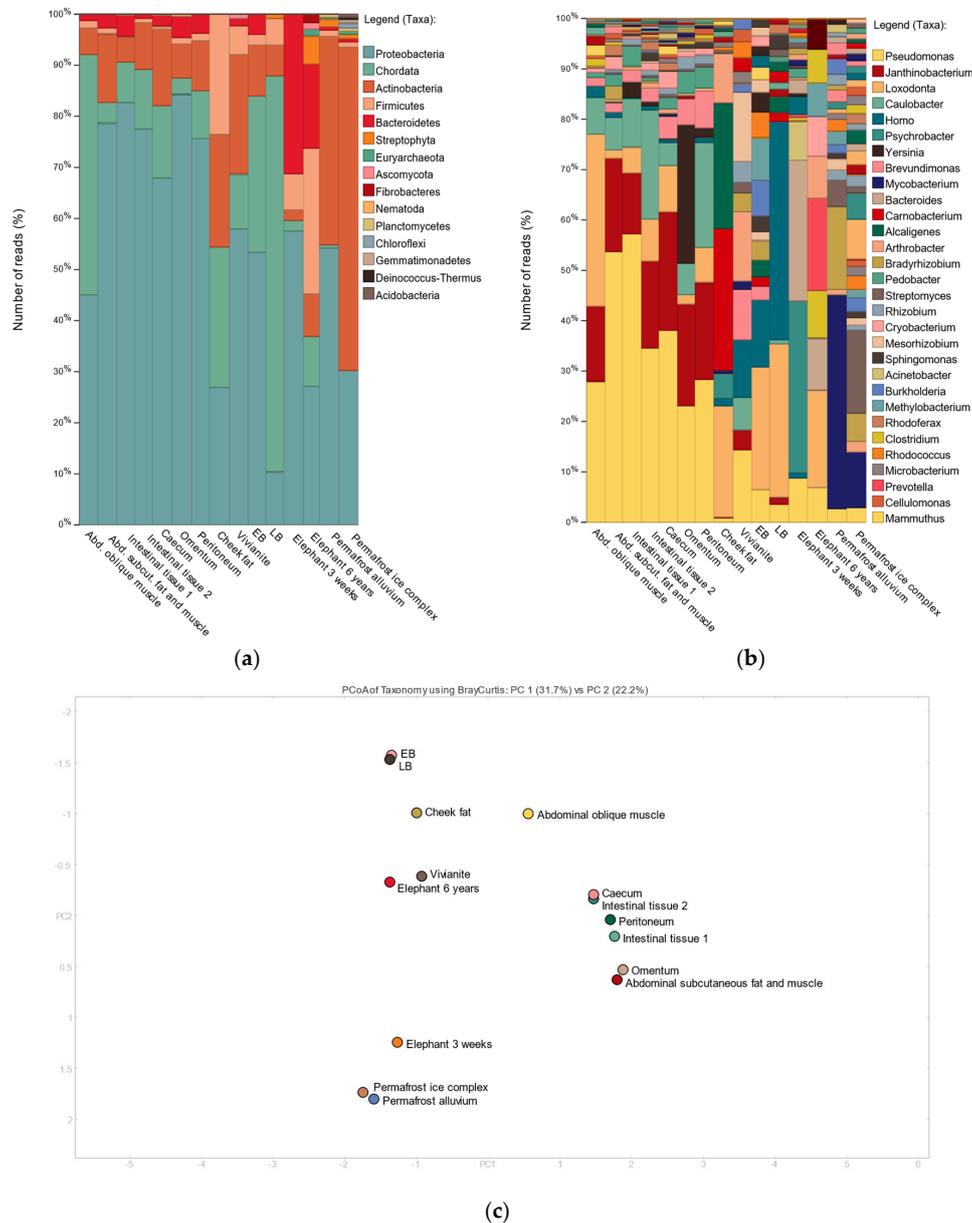
Figure 3. Taxonomic composition of tissue-specific metagenomes. Quality-filtered, deduplicated data were compared to the National Center for Biotechnology Information (NCBI) nucleotide collection with megablast and results were visualized with MEGAN6. Taxa are displayed at the phylum (**a**) and genus (**b**) level. The presence of *Yersinia* in the omentum sample and of *Carnobacterium* and *Alcaligenes* in the cheek fat sample characterize these tissues. For visualization purposes only the most abundant taxa are listed. (**c**) The Bray-Curtis Principal Coordinate Analysis (PCoA) of taxonomic profiles ranked by species shows samples to differ in their taxonomic composition from the non-template controls and the vivianite environmental control, as well as from the permafrost and elephant fecal samples. EB = extraction blank; LB = library blank.

3.3. 16S Metagenomic Analysis

The 16S metagenomic analysis retained between 9 (library blank) and 8037 (elephant 3 weeks) unique sequences (mean: 1739 sequences per sample; standard deviation: 1988 sequences) after clustering to the SILVA database. We identified *Yersinia* in the omentum sample at a 0.84% relative frequency and *Alcaligenes* and *Carnobacterium* in the cheek fat sample at 1.77% and 1.56% relative frequencies, respectively (Figure S4). All PCoA analyses produced concordant results: With the exception of the cheek fat sample, the Lyuba tissues clustered with themselves near the controls (Figure 4a and Figure S5). The cheek fat sample was the most similar to the elephant feces. The tissue samples did not cluster with the permafrost sample. These results are concordant with those of the SourceTracker analysis (Figure 4b): ~10% of tissue reads derived from the local environment (represented by vivianite) and ~5% derive from laboratory contamination. The cheek fat sample had the lowest laboratory contamination level (2%) which may explain its differentiation in the PCoAs. Interestingly, the omentum had the highest proportion of reads that correspond to the elephant fecal microbiome (11%), but still clustered with the other Lyuba tissue samples.

(**a**)
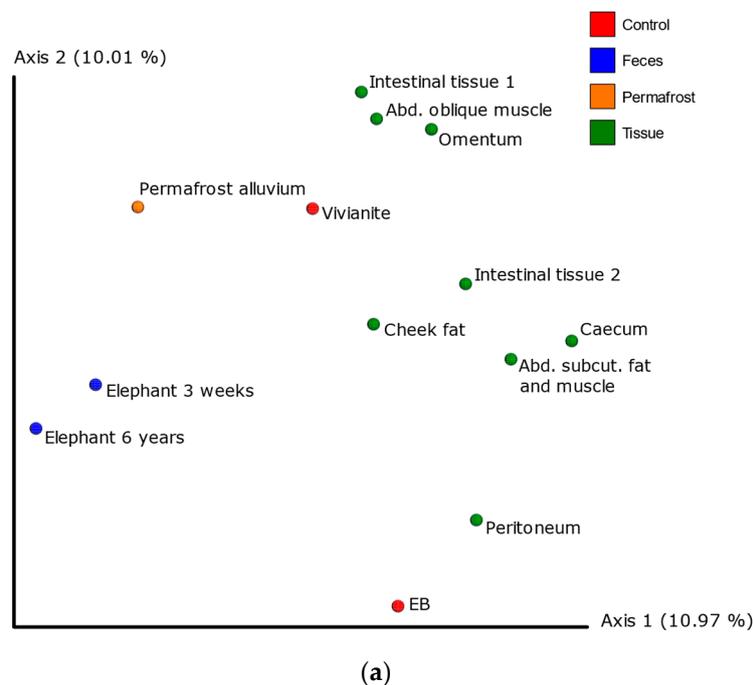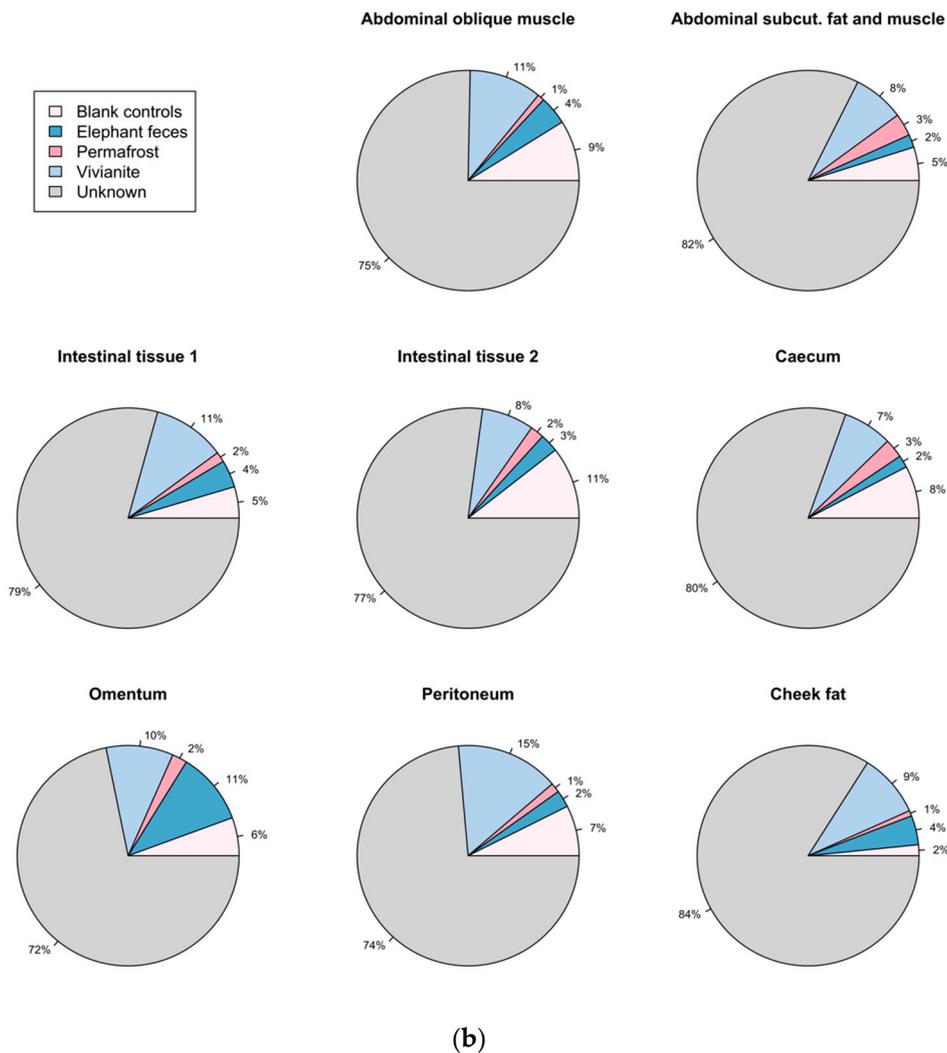
Figure 4. *Cont.*

Figure 4. QIIME2 16S metagenomic analysis. (**a**) PCoA of Unweighted UniFrac distances. Phylogenetic diversity metrics were calculated with rarefication to 149 sequences. (**b**) SourceTracker analysis. Laboratory blank controls, the vivianite environmental control, the elephant fecal samples, and the Russian permafrost samples were set as possible sources of microbial communities.

3.4. MetaPhlAn2 Analysis

The MetaPhlAn2 analysis results largely replicated the results of the megablast and MALT analyses (Figure S6). With the exception of the cheek fat sample, the Lyuba tissue metagenomes were dominated by *Pseudomonas*, *Janthinobacterium*, *Caulobacter*, *Pedobacter*, and *Brevundimonas*. *Yersinia intermedia* was detected in the omentum and intestinal tissue 1 sample. *Alcaligenes*, *Carnobacterium* sp. 17.4, and Arthrobacter gangostriensis were found in the cheek fat. Only *Brevundimonas* was identified in the vivianite. With the exception of the cheek fat, the Lyuba tissue samples cluster together as in the other analyses.

3.5. Authentication of Potential Ancient Bacterial Signals

While the taxonomic composition of Lyuba's tissue-specific metagenomes was overall similar, the presence of *Yersinia*, *Carnobacterium* sp. CP1, and *A. faecalis* as prominent components in the omentum and cheek fat samples' metagenomes did set these two tissues apart. In order to test whether the presence of these bacteria represented authentic ancient host-associated microbiomic signatures, we mapped all reads obtained from the omentum sample against the reference genomes of various

Yersinia species, and the reads obtained from the cheek fat sample to the *A. faecalis* reference genome and the *Carnobacterium* sp. CP1 genome. We then analyzed the DNA damage patterns of all mapped reads to test whether we could detect typical aDNA deamination.

For the omentum sample we obtained the most reads when mapping against the reference genome of *Y. intermedia* (751,425 reads mapping with a quality of 25 or higher, 10.9% of unique reads). While *Y. intermedia* has been associated with opportunistic gastrointestinal infections, this bacterium is widely found in the environment, mostly in fresh water [61]. Furthermore, a mapDamage analysis showed no increase of cytosine deamination rates at 5′-overhangs (Figure 5), which is inconsistent with authentic aDNA [42].

Similarly, while we obtained 519,810 reads (4.7% of unique reads) mapping to the *Carnobacterium* reference genome and 549,975 reads (5% of unique reads) mapping to the *A. faecalis* reference genome (with a quality of 25 or higher) we failed to detect DNA damage patterns consistent with aDNA (Figure 5).
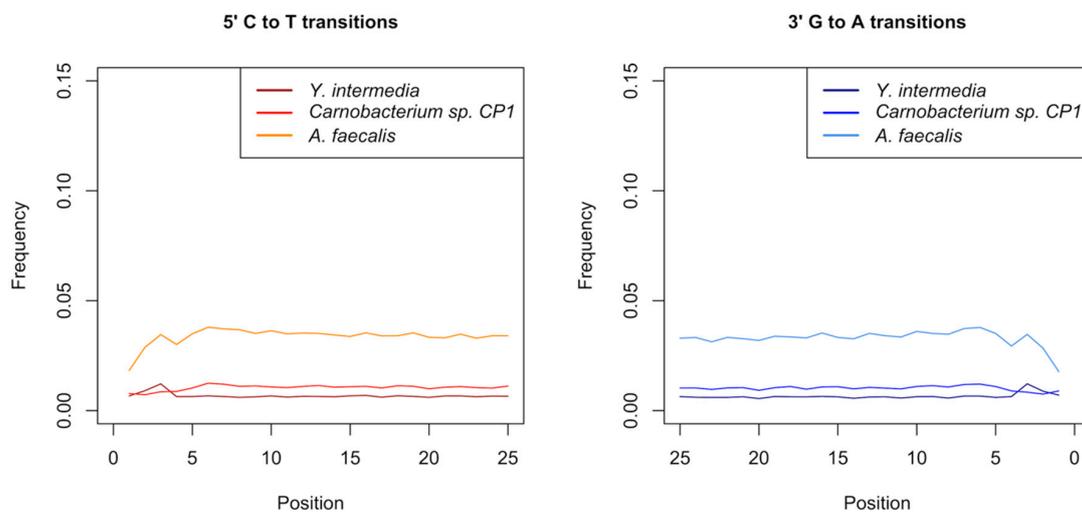
Figure 5. DNA damage pattern analysis of omentum reads mapped to *Y. intermedia* and cheek fat reads mapped to *Carnobacterium* sp. CP1 and *Alcaligenes faecalis*. Quality-filtered reads were mapped to the reference genomes with a minimum mapping quality of 25 and DNA damage plots were generated using mapDamage. No increase in cytosine deamination rates at 5′-overhangs is visible (C to T and G to A transitions), which is inconsistent with aDNA.

4. Discussion

Ancient microbiome studies are of great value in the context of human health and evolution. However, microbial communities reconstructed from archaeological material typically have a large environmental component. This can lead to false positive signals [14,15] and strict validation procedures are necessary. Additionally, the introduction of microbial contaminants via laboratory reagents also needs to be considered [18–20]. Excellent reviews and guidelines on these issues have already been provided by others (e.g., References [16,17]). Based on these guidelines, we attempted to reconstruct the tissue-specific metagenomes of Lyuba, a 42,000-year-old, permafrost-preserved woolly mammoth calf, together with environmental and non-template experimental controls and compared them to metagenomes obtained from permafrost and elephant fecal samples. The endogenous mammoth DNA content was very variable across samples (2–43%), but we obtained DNA deamination patterns consistent with authentic aDNA for all tissue samples [42]. Furthermore, the mitogenome consensus sequence we obtained was consistent with the previously published Lyuba mitogenome [44].

A preliminary taxonomic composition analysis suggested that, despite a large component derived from environmental or laboratory contamination, metagenomes reconstructed from Lyuba's tissue samples were distinct from those obtained from controls. We paid particular attention to the presence

of *Yersinia* in the omentum sample, and of *A. faecalis* and *Carnobacterium* sp. CP1 in the cheek fat sample. These taxa were absent or nearly absent from all other tissues and from laboratory and environmental controls. Therefore, the detection of several hundred thousand of reads in one particular tissue in an otherwise rather uniform taxonomic background, led us to think that these could represent ancient microbiomic signals originating, for example, from an infection in Lyuba's tissues. Further corroborating this hypothesis, several *Yersinia* species are known pathogens or are at least capable of opportunistic infections [61]. *A. faecalis* is an opportunistic pathogen as well [62] and while there are no known pathogens in the genus *Carnobacterium* [63], the presence of this lactic-acid-producing bacterium is consistent with the proposed explanation that acidification of Lyuba's tissues contributed to its exceptional preservation [25].

However, these hypotheses were not corroborated by aDNA authentication criteria. In order to identify which *Yersinia* species was most likely to be present in the omentum sample, we mapped all reads against several *Yersinia* reference genomes and obtained the most reads for *Y. intermedia*, a species that is widely distributed in the environment [61]. Additionally, we analyzed deamination rates at 5′-overhangs of omentum reads mapped to *Y. intermedia*, as well as of cheek fat reads mapped to *A. faecalis* and *Carnobacterium* sp. CP1 and found no increase of cytosine deamination, which is inconsistent with authentic aDNA [42]. For these reasons, we cannot conclude that the presence of these bacteria in the omentum and cheek fat samples represents authentic ancient microbiomic signals. We believe that these signals are much more likely to have originated from contamination instead.

We observed several bacterial taxa that were present in almost all of Lyuba's tissues as well as in one or more of the laboratory and vivianite environmental controls (e.g., *Pseudomonas*, *Janthinobacterium*, *Caulobacter*, and *Brevundimonas*). In this case, the source of contamination can easily be attributed to the environment or to laboratory reagents or workflows. The 16S metagenomic and SourceTracker analyses indicate that our results are affected by these contaminants. However, *Yersinia*, *A. faecalis*, and *Carnobacterium* sp. CP1 were unique to the omentum and cheek fat samples, respectively, with the exception of very few reads that can be explained by index switching. While environmental or laboratory contamination cannot be excluded, they are less likely. Other possible sources of contamination that consider the tissue specificity we observed are contamination during the sampling procedure or a non-uniform colonization of Lyuba's tissues by environmental bacteria. Lyuba likely endured at least one cycle of thawing and freezing before discovery [25,27]. This may have affected DNA recovery and the fact that we observed very variable amounts of endogenous DNA in the different tissues could reflect variations in DNA preservation. Furthermore, thawing may have offered an opportunity for colonization of Lyuba's tissues by environmental microorganisms, and if this process did not take place in a homogenous way, it could explain different contaminants among tissues. Moreover, part of the investigated tissues, such as fat and muscle, should be free of bacterial signals, which also indicates post-mortem contamination as a source of the detected bacterial signals.

We conclude that, even in ideal morphological preservation circumstances, host-endogenous microbiome signals can be swamped out by contaminating signals. Therefore, microbiome analyses from ancient tissues must be done with the utmost care, because even when environmental and laboratory controls are not compositionally similar to experimental samples, this does not indicate that microbes are necessarily endogenous.

Author Contributions: Conceptualization, N.B., A.P. and F.R.; Data curation, G.F. and M.G.C.; Formal analysis, G.F., H.E.L.L., J.N. and M.G.C.; Investigation, A.S.B. and M.G.C.; Methodology, G.F., A.S.B. and M.G.C.; Project administration, F.R.; Resources, H.E.L.L., N.B., A.P., F.R. and M.C.; Supervision, F.R., A.S.B. and M.G.C.; Visualization, G.F. and J.N.; Writing—original draft, G.F., E.R. and M.G.C.; Writing—review & editing, G.F., J.N., A.S.B. and M.G.C.

References

1. Paabo, S. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* **1985**, *314*, 644–645. [CrossRef] [PubMed]

2. Taylor, G.M.; Crossey, M.; Saldanha, J.; Waldron, T. DNA from *Mycobacterium tuberculosis* identified in Mediaeval human skeletal remains using polymerase chain reaction. *J. Archaeol. Sci.* **1996**, *23*, 789–798. [CrossRef]

3. Nerlich, A.G.; Haas, C.J.; Zink, A.; Szeimies, U.; Hagedorn, H.G. Molecular evidence for tuberculosis in an ancient Egyptian mummy. *Lancet* **1997**, *350*, 1404. [CrossRef]

4. Zink, A.; Reischl, U.; Wolf, H.; Nerlich, A.G. Molecular evidence of bacteremia by gastrointestinal pathogenic bacteria in an infant mummy from ancient Egypt. *Arch. Pathol. Lab. Med.* **2000**, *124*, 1614–1618. [PubMed]

5. Bos, K.I.; Schuenemann, V.J.; Golding, G.B.; Burbano, H.A.; Waglechner, N.; Coombes, B.K.; McPhee, J.B.; DeWitte, S.N.; Meyer, M.; Schmedes, S.; et al. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **2011**, *478*, 506–510. [CrossRef] [PubMed]

6. Schuenemann, V.J.; Singh, P.; Mendum, T.A.; Krause-Kyora, B.; Jager, G.; Bos, K.I.; Herbig, A.; Economou, C.; Benjak, A.; Busso, P.; et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **2013**, *341*, 179–183. [CrossRef] [PubMed]

7. Adler, C.J.; Dobney, K.; Weyrich, L.S.; Kaidonis, J.; Walker, A.W.; Haak, W.; Bradshaw, C.J.A.; Townsend, G.; Sołtysiak, A.; Alt, K.W.; et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **2013**, *45*, 450–455. [CrossRef] [PubMed]

8. Warinner, C.; Rodrigues, J.F.M.; Vyas, R.; Trachsel, C.; Shved, N.; Grossmann, J.; Radini, A.; Hancock, Y.; Tito, R.Y.; Fiddyment, S.; et al. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **2014**, *46*, 336–344. [CrossRef] [PubMed]

9. Hooper, L.V.; Gordon, J.I. Commensal host-bacterial relationships in the gut. *Science* **2001**, *292*, 1115–1118. [CrossRef] [PubMed]

10. Tremaroli, V.; Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **2012**, *489*, 242–249. [CrossRef] [PubMed]

11. Yatsunenko, T.; Rey, F.E.; Manary, M.J.; Trehan, I.; Dominguez-Bello, M.G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R.N.; Anokhin, A.P.; et al. Human gut microbiome viewed across age and geography. *Nature* **2012**, *486*, 222. [CrossRef] [PubMed]

12. Tito, R.Y.; Knights, D.; Metcalf, J.; Obregon-Tito, A.J.; Cleeland, L.; Najar, F.; Roe, B.; Reinhard, K.; Sobolik, K.; Belknap, S.; et al. Insights from characterizing extinct human gut microbiomes. *PLoS ONE* **2012**, *7*, e51146. [CrossRef] [PubMed]

13. Weyrich, L.S.; Duchene, S.; Soubrier, J.; Arriola, L.; Llamas, B.; Breen, J.; Morris, A.G.; Alt, K.W.; Caramelli, D.; Dresely, V.; et al. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **2017**, *544*, 357–361. [CrossRef] [PubMed]

14. Campana, M.G.; Robles Garcia, N.; Ruhli, F.J.; Tuross, N. False positives complicate ancient pathogen identifications using high-throughput shotgun sequencing. *BMC Res. Notes* **2014**, *7*, 111. [CrossRef] [PubMed]

15. Bos, K.I.; Jager, G.; Schuenemann, V.J.; Vagene, A.J.; Spyrou, M.A.; Herbig, A.; Nieselt, K.; Krause, J. Parallel detection of ancient pathogens via array-based DNA capture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20130375. [CrossRef] [PubMed]

16. Warinner, C.; Herbig, A.; Mann, A.; Fellows Yates, J.A.; Weiß, C.L.; Burbano, H.A.; Orlando, L.; Krause, J. A robust framework for microbial archaeology. *Annu. Rev. Genom. Hum. Genet.* **2017**, *18*, 321–356. [CrossRef] [PubMed]

17. Key, F.M.; Posth, C.; Krause, J.; Herbig, A.; Bos, K.I. Mining metagenomic data sets for ancient DNA: Recommended protocols for authentication. *Trends Genet.* **2017**, *33*, 508–520. [CrossRef] [PubMed]

18. Salter, S.J.; Cox, M.J.; Turek, E.M.; Calus, S.T.; Cookson, W.O.; Moffatt, M.F.; Turner, P.; Parkhill, J.; Loman, N.J.; Walker, A.W. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **2014**, *12*. [CrossRef] [PubMed]

19. Glassing, A.; Dowd, S.E.; Galandiuk, S.; Davis, B.; Chiodini, R.J. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* **2016**, *8*, 24. [CrossRef] [PubMed]

20. Lauder, A.P.; Roche, A.M.; Sherrill-Mix, S.; Bailey, A.; Laughlin, A.L.; Bittinger, K.; Leite, R.; Elovitz, M.A.; Parry, S.; Bushman, F.D. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome* **2016**, *4*, 29. [CrossRef] [PubMed]

21. Rohland, N.; Malaspinas, A.-S.; Pollack, J.L.; Slatkin, M.; Matheus, P.; Hofreiter, M. Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol.* **2007**, *5*, e207. [CrossRef] [PubMed]

22. Long, S.Y.; Latimer, E.M.; Hayward, G.S. Review of elephant endotheliotropic Herpesviruses and acute hemorrhagic disease. *ILAR J.* **2015**, *56*, 283–296. [CrossRef] [PubMed]

23. Zlot, A.; Vines, J.; Nystrom, L.; Lane, L.; Behm, H.; Denny, J.; Finnegan, M.; Hostetler, T.; Matthews, G.; Storms, T.; et al. Diagnosis of tuberculosis in three zoo elephants and a human contact—Oregon, 2013. *Centers Dis. Control Prev. Morb. Mortal. Wkly. Rep.* **2016**, *64*, 1398–1402. [CrossRef] [PubMed]

24. Kosintsev, P.A.; Lapteva, E.G.; Trofimova, S.S.; Zanina, O.G.; Tikhonov, A.N.; van der Plicht, J. The intestinal contents of a baby woolly mammoth (*Mammuthus primigenius* Blumenbach, 1799) from the Yuribey River (Yamal Peninsula). *Dokl. Biol. Sci.* **2010**, *432*, 209–211. [CrossRef] [PubMed]

25. Fisher, D.C.; Tikhonov, A.N.; Kosintsev, P.A.; Rountrey, A.N.; Buigues, B.; van der Plicht, J. Anatomy, death, and preservation of a woolly mammoth (*Mammuthus primigenius*) calf, Yamal Peninsula, northwest Siberia. *Quat. Int.* **2012**, *255*, 94–105. [CrossRef]

26. Rountrey, A.N.; Fisher, D.C.; Tikhonov, A.N.; Kosintsev, P.A.; Lazarev, P.A.; Boeskorov, G.; Buigues, B. Early tooth development, gestation, and season of birth in mammoths. *Quat. Int.* **2012**, *255*, 196–205. [CrossRef]

27. Papageorgopoulou, C.; Link, K.; Rühli, F.J. Histology of a woolly mammoth (*Mammuthus primigenius*) preserved in permafrost, Yamal Peninsula, Northwest Siberia. *Anat. Rec.* **2015**, *298*, 1059–1071. [CrossRef] [PubMed]

28. Krüttli, A.; Bouwman, A.; Akgül, G.; Della Casa, P.; Rühli, F.; Warinner, C. Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in medieval central Europe. *PLoS ONE* **2014**, *9*, e86251. [CrossRef] [PubMed]

29. Cooper, A.; Poinar, H.N. Ancient DNA: Do it right or not at all. *Science* **2000**, *289*, 1139. [CrossRef] [PubMed]

30. Llamas, B.; Valverde, G.; Fehren-Schmitz, L.; Weyrich, L.S.; Cooper, A.; Haak, W. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR Sci. Technol. Archaeol. Res.* **2017**, *3*, 1–14. [CrossRef]

31. Meyer, M.; Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, *2010*. [CrossRef] [PubMed]

32. Kircher, M.; Sawyer, S.; Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **2012**, *40*, e3. [CrossRef] [PubMed]

33. Renaud, G.; Stenzel, U.; Maricic, T.; Wiebe, V.; Kelso, J. deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **2015**, *31*, 770–772. [CrossRef] [PubMed]

34. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

35. SeqPrep. Available online: https://github.com/jstjohn/SeqPrep (accessed on 19 February 2016).

36. Maixner, F.; Krause-Kyora, B.; Turaev, D.; Herbig, A.; Hoopmann, M.R.; Hallows, J.L.; Kusebauch, U.; Vigl, E.E.; Malfertheiner, P.; Megraud, F.; et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* **2016**, *351*, 162–165. [CrossRef] [PubMed]

37. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 8 April 2014).

38. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]

39. Schubert, M.; Ginolhac, A.; Lindgreen, S.; Thompson, J.F.; AL-Rasheid, K.A.; Willerslev, E.; Krogh, A.; Orlando, L. Improving ancient DNA read mapping against modern reference genomes. *BMC Genom.* **2012**, *13*, 178. [CrossRef] [PubMed]

40. Picard. Available online: http://broadinstitute.github.io/picard/ (accessed on 5 February 2016).

41. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

42. Jónsson, H.; Ginolhac, A.; Schubert, M.; Johnson, P.; Orlando, L. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **2013**, *29*, 1682–1684. [CrossRef] [PubMed]

43. Harris, R.S. Improved Pairwise Alignment of Genomic DNA. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2007.

44. Enk, J.; Devault, A.; Widga, C.; Saunders, J.; Szpak, P.; Southon, J.; Rouillard, J.-M.; Shapiro, B.; Golding, G.B.; Zazula, G.; et al. *Mammuthus* population dynamics in Late Pleistocene North America: Divergence, phylogeography and introgression. *Front. Ecol. Evol.* **2016**, *4*. [CrossRef]

45. Ilmberger, N.; Güllert, S.; Dannenberg, J.; Rabausch, U.; Torres, J.; Wemheuer, B.; Alawi, M.; Poehlein, A.; Chow, J.; Turaev, D.; et al. A comparative metagenome survey of the fecal microbiota of a breast- and a plant-fed Asian elephant reveals an unexpectedly high diversity of glycoside hydrolase family enzymes. *PLoS ONE* **2014**, *9*, e106707. [CrossRef] [PubMed]

46. Krivushin, K.; Kondrashov, F.; Shmakova, L.; Tutukina, M.; Petrovskaya, L.; Rivkina, E. Two metagenomes from late Pleistocene northeast Siberian permafrost. *Genome Announc.* **2015**, *3*, e01380-14. [CrossRef] [PubMed]

47. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]

48. Zhang, Z.; Schwartz, S.; Wagner, L.; Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **2000**, *7*, 203–214. [CrossRef] [PubMed]

49. Vågene, Å.J.; Herbig, A.; Campana, M.G.; Robles García, N.M.; Warinner, C.; Sabin, S.; Spyrou, M.A.; Andrades Valtueña, A.; Huson, D.; Tuross, N.; et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2018**, *2*, 520–528. [CrossRef] [PubMed]

50. Huson, D.H.; Beier, S.; Flade, I.; Górska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.-J.; Tappu, R. MEGAN Community edition—interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **2016**, *12*, e1004957. [CrossRef] [PubMed]

51. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335–336. [CrossRef] [PubMed]

52. Pruesse, E.; Quast, C.; Knittel, K.; Fuchs, B.M.; Ludwig, W.; Peplies, J.; Glöckner, F.O. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **2007**, *35*, 7188–7196. [CrossRef] [PubMed]

53. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [CrossRef] [PubMed]

54. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]

55. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef] [PubMed]

56. Vázquez-Baeza, Y.; Pirrung, M.; Gonzalez, A.; Knight, R. EMPeror: A tool for visualizing high-throughput microbial community data. *Gigascience* **2013**, *2*, 16. [CrossRef] [PubMed]

57. Vázquez-Baeza, Y.; Gonzalez, A.; Smarr, L.; McDonald, D.; Morton, J.T.; Navas-Molina, J.A.; Knight, R. Bringing the dynamic microbiome to life with animations. *Cell Host Microbe* **2017**, *21*, 7–10. [CrossRef] [PubMed]

58. Knights, D.; Kuczynski, J.; Charlson, E.S.; Zaneveld, J.; Mozer, M.C.; Collman, R.G.; Bushman, F.D.; Knight, R.; Kelley, S.T. Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* **2011**, *8*, 761–763. [CrossRef] [PubMed]

59. Truong, D.T.; Franzosa, E.A.; Tickle, T.L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **2015**, *12*, 902–903. [CrossRef] [PubMed]

60. Dabney, J.; Meyer, M.; Pääbo, S. Ancient DNA Damage. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a012567. [CrossRef] [PubMed]

61. Sulakvelidze, A. Yersiniae other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: The ignored species. *Microbes Infect.* **2000**, *2*, 497–513. [CrossRef]

62. Phung, L.T.; Trimble, W.L.; Meyer, F.; Gilbert, J.A.; Silver, S. Draft genome sequence of *Alcaligenes faecalis* subsp. *faecalis* NCIB 8687 (CCUG 2071). *J. Bacteriol.* **2012**, *194*, 5153. [CrossRef] [PubMed]

63. Leisner, J.J.; Laursen, B.G.; Prévost, H.; Drider, D.; Dalgaard, P. Carnobacterium: Positive and negative effects in the environment and in foods. *FEMS Microbiol. Rev.* **2007**, *31*, 592–613. [CrossRef] [PubMed]