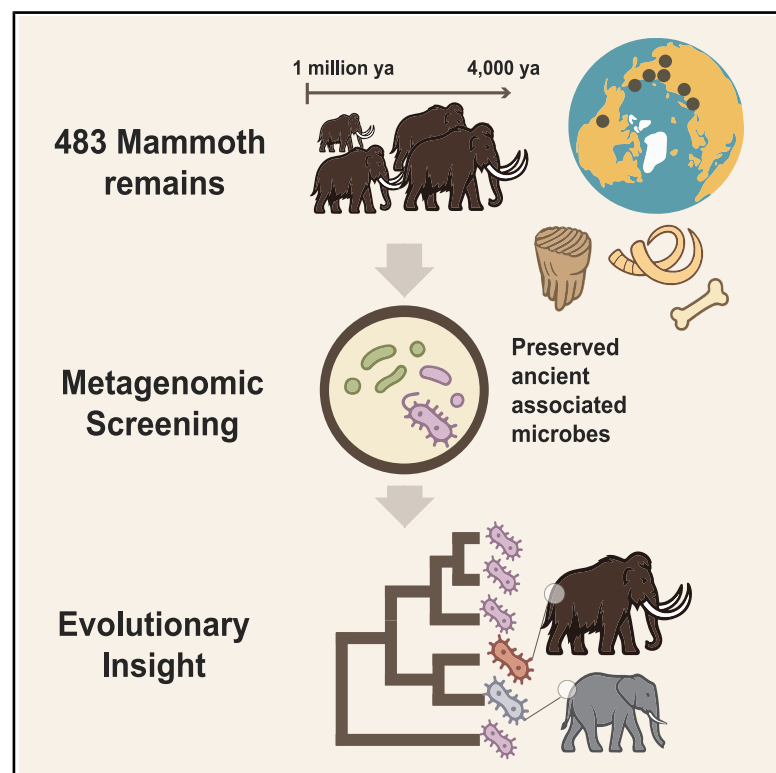


Ancient host-associated microbes obtained from mammoth remains

Graphical abstract



Authors

Benjamin Guinet, Nikolay Oskolkov, Kelsey Moreland, ..., Pavel Nikolskiy, Love Dalén, Tom van der Valk

Correspondence

benjamin.guinet95@gmail.com

In brief

A study of 483 mammoth remains spanning over a million years reveals ancient host-associated microbes preserved in bones and teeth, offering new insights into the long-term evolution of Pleistocene microbiomes and their association with megafaunal hosts.

Highlights

- Analysis of microbial DNA in 483 mammoths, dated from >1 million years ago to near-extinction
- Six microbial clades persisted across diverse regions and extended time periods
- Partial *Erysipelothrix* genome recovered from a 1.1-million-year-old steppe mammoth

Article

Ancient host-associated microbes obtained from mammoth remains

Benjamin Guinet,^{1,2,26,*} Nikolay Oskolkov,³ Kelsey Moreland,¹ Marianne Dehasque,^{1,2,4} J. Camilo Chacón-Duque,^{1,4,5} Anders Angerbjörn,⁴ Juan Luis Arsuaga,^{20,21} Gleb Danilov,¹² Foteini Kanellidou,^{1,4} Andrew C. Kitchener,^{23,24} Héloïse Muller,¹¹ Valerii Plotnikov,⁹ Albert Protopopov,⁹ Alexei Tikhonov,^{10,19} Laura Termes,¹⁷ Grant Zazula,^{14,15} Peter Mortensen,²⁵ Lena Grigorieva,¹⁸ Michael Richards,¹⁷ Beth Shapiro,¹⁶ Adrian M. Lister,¹³ Sergey Vartanyan,⁸ David Díez-del-Molino,^{1,2,4} Anders Götherström,^{1,5} Patrícia Pečnerová,^{2,4,7} Pavel Nikolskiy,⁶ Love Dalén,^{1,2,4} and Tom van der Valk^{1,2,22}

¹Centre for Palaeogenetics, Svante Arrhenius väg 20C, 10691 Stockholm, Sweden

²Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

³Department of Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Lund University, Lund, Sweden

⁴Department of Zoology, Stockholm University, 10691 Stockholm, Sweden

⁵Department of Archaeology and Classical Studies, Stockholm University, Lilla Frescativägen 7, 11418 Stockholm, Sweden

⁶Geological Institute of the Russian Academy of Sciences, Moscow, Russia

⁷Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark

⁸North-East Interdisciplinary Scientific Research Institute N.A. Shilo, Far East Branch, Russian Academy of Sciences, Magadan, Russia

⁹Academy of Sciences of Sakha Republic, Lenin Avenue 33, Yakutsk, Republic of Sakha (Yakutia), Russia

¹⁰Zoological Institute of Russian Academy of Sciences, Saint-Petersburg, Russia

¹¹Master de Biologie, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon I, Université de Lyon, 69007 Lyon, France

¹²Peter the Great Museum of Anthropology and Ethnography, Kunstkamera, Russian Academy of Sciences, 3 University Embankment, Box 199034, Saint-Petersburg, Russia

¹³Natural History Museum, London SW7 5BD, UK

¹⁴Yukon Government, Palaeontology Program, Department of Tourism and Culture, Whitehorse, YT, Canada

¹⁵Collections and Research, Canadian Museum of Nature, Ottawa, ON, Canada

¹⁶Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, CA, USA

¹⁷Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

¹⁸Center of Molecular Paleontology, M.K. Ammosov North-Eastern Federal University, Yakutsk 677000, Russia

¹⁹Museum of Mammoth, Institute of the Applied Ecology of the North, North-Eastern Federal University, Yakutsk 677000, Russia

²⁰Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, Madrid, Spain

²¹Departamento de Geodinámica, Estratigrafía y Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, Madrid, Spain

²²Science for life Laboratory, 17165 Stockholm, Sweden

²³Department of Natural Sciences, National Museums Scotland, Chambers Street, Edinburgh EH1 1JF, UK

²⁴School of Geosciences, University of Edinburgh, Drummond Street, Edinburgh EH8 9XP, UK

²⁵Department of Zoology, Swedish Museum of Natural History, Stockholm, Sweden

²⁶Lead contact

*Correspondence: benjamin.guinet95@gmail.com

<https://doi.org/10.1016/j.cell.2025.08.003>

SUMMARY

Ancient genomic studies have extensively explored human-microbial interactions, yet research on non-human animals remains limited. In this study, we analyzed ancient microbial DNA from 483 mammoth remains spanning over 1 million years, including 440 newly sequenced and unpublished samples from a 1.1-million-year-old steppe mammoth. Using metagenomic screening, contaminant filtering, damage pattern analysis, and phylogenetic inference, we identified 310 microbes associated with different mammoth tissues. While most microbes were environmental or post-mortem colonizers, we recovered genomic evidence of six host-associated microbial clades spanning *Actinobacillus*, *Pasteurella*, *Streptococcus*, and *Erysipelothrix*. Some of these clades contained putative virulence factors, including a *Pasteurella*-related bacterium that had previously been linked to the deaths of African elephants. Notably, we reconstructed partial genomes of *Erysipelothrix* from the oldest mammoth sample, representing the oldest authenticated host-associated microbial DNA to date. This work demonstrates the potential of obtaining ancient animal microbiomes, which can inform further paleoecological and evolutionary research.

INTRODUCTION

The sequencing of mammoth (*Mammuthus*) DNA has enabled comprehensive studies on mammoth evolution, biogeography, and ecology.^{1–5} However, past interactions between microbes and extinct megafauna remain largely unexplored. Investigating these relationships could provide insights into the role of microbes in adaptation to extreme environments, the impact of population size fluctuations on the microbiome during glacial and interglacial periods, dietary shifts over time, and the potential role of microbes in megafaunal extinction. For example, Asian elephants, *Elephas maximus*, the closest living relatives to mammoths, suffer from pathogens including a virus with high mortality in calves (endotherotropic herpesvirus) and a bacterium causing anthrax disease (*Bacillus anthracis*),^{6–9} prompting consideration of whether similar microbes were affecting their extinct mammoth cousins. Ancient remains such as teeth and bones can preserve not only the host's DNA, but also the DNA of microbes that co-occurred at the individual's time of death.¹⁰ These data have now emerged as a valuable resource for understanding pandemics, lifestyle patterns, and population dynamics.^{11–13} Here, we aimed to explore past interaction between mammoths and co-occurring microbes along a time-span from over 1 million years ago until the extinction of mammoths on Wrangel Island 4,000 years ago. We analyzed a total of 483 genomic datasets generated from various tissues including teeth, molars, skulls, and skin tissue of mammoths, of which 440 are newly sequenced and unpublished samples, including new sequence data obtained from a 1.1-million-year-old steppe mammoth (*Mammuthus trogontherii*) sample (Figure S1; Table S1).

RESULTS

Microbial screening of mammoth sequence data

To screen the mammoth samples for the presence of microbial ancient DNA (aDNA), we first built a *Kraken2* microbial database, containing complete genomes for over 500,000 genomes (see the method explanation in the STAR Methods and the full pipeline abstract in Figure S2). We used the genome taxonomy database (GTDB) as it is more comprehensive than the pre-built *Kraken2* microbial RefSeq database, resulting in a higher number of classified reads (see detail in Data S1, section A). Next, we collected all published genome data for mammoths, as well as in-house mammoth sequence data that has been generated over the past decade. We removed from each of the mammoth samples the putative mammoth sequences as well as contaminant human reads that might have been introduced during sample processing by aligning all the data to a concatenated reference containing the Asian elephant, human, and mammoth mitogenome assemblies.¹⁴ Subsequently, all reads that did not align were classified against the microbial database using a k-mer approach (see STAR Methods, metagenomic screening for more details; Table S2). Across all samples, between 0.25% and 68.5% of the reads could be classified against this database (mean = 31.1%) (Table S2). To filter for bacterial contamination that may have been introduced during laboratory sample processing, as well as for non-host-associated microbes that colonized the mammoth tissues post-mortem, we analyzed 43 laboratory blanks collected in the laboratory at

the same time as the mammoth samples, as well as 22 previously sequenced samples from different ancient Arctic sediments.^{16,17} Using a k-mer based genetic distance approach,¹⁵ we find that most mammoth samples exhibited similar read compositions, with only a small subset of samples also sharing high genetic similarities to the sediment samples (see detail in Data S1, section B). All laboratory blanks displayed a distinctly different read composition, suggesting little laboratory contamination (Data S1, section B). Between 1.8% and 72.3% of the microbial species identified among mammoths were also classified in at least one laboratory blank or Arctic sediment sample and thus excluded from downstream analysis. Next, to align the remaining unmapped candidate microbial reads, we built a *bowtie2* index, comprising all microbial species with at least 200 reads classified in the *Kraken2* analysis ($n = 87,958$), taking one representative species for each genus in cases multiple species within a genus were detected. Using *bowtie2*, we then competitively mapped the reads against the genomes of all previously classified microbial species. Among all the reference microbial genomes in the alignments, 28.9% exhibited an evenness of coverage over 90% (uniformness with which aligned reads are distributed across the entire reference genome) with an average coverage of 0.82X (one-tailed std = 3.83X) (Figure 1B), and these were targeted for further investigation.

Description and authentication of ancient microbes

Most of the mammoth samples in this study were subjected to enzymatic uracil-DNA glycosylase (UDG) treatment, a process that eliminates typical patterns of ancient DNA damage, except at methylated sites in vertebrates.¹⁹ In vertebrates DNA methylation occurs at the 5'-position of cytosine residues of CpG dinucleotides in somatic tissues. This methylation inhibits the effect of the UDG treatment at these positions, preserving the C-to-T damage pattern at CpG sites.¹⁹ However, the mechanisms of cytosine methylation in prokaryotes remain poorly understood, and it is unclear whether these mechanisms are consistent across all bacterial clades.²⁰ As a result, this enzymatic UDG treatment limited our ability to detect typical ancient DNA damage patterns (Figure 1E). Therefore, we considered deamination at CpG sites in cases where bacteria exhibited an ancient DNA damage pattern at these sites, but we predominantly relied on the observation that UDG treatment does not always completely remove C-to-T damage, leaving residual C-to-T damage at the first three bases of the reads, both in the mammoth and in the microbial reads (Figures 2A and 2C; Table S3).

To further authenticate the candidate ancient microbes, we constructed a quality control scoring pipeline that relied on several strict filters, including deamination patterns at remaining C-to-T damage sites, evenness of coverage, and average read lengths, which are expected to be short due to post-mortem DNA fragmentation (see STAR Methods, authentication of ancient DNA for score metric details). Of the microbes identified in the laboratory blanks, only 0.04% got the highest score of five (Figure 1F), whereas 31.82% of the ancient sequence data from Arctic sediments and 0.51% of the microbes identified in the mammoth samples were scored as ancient. We then visually

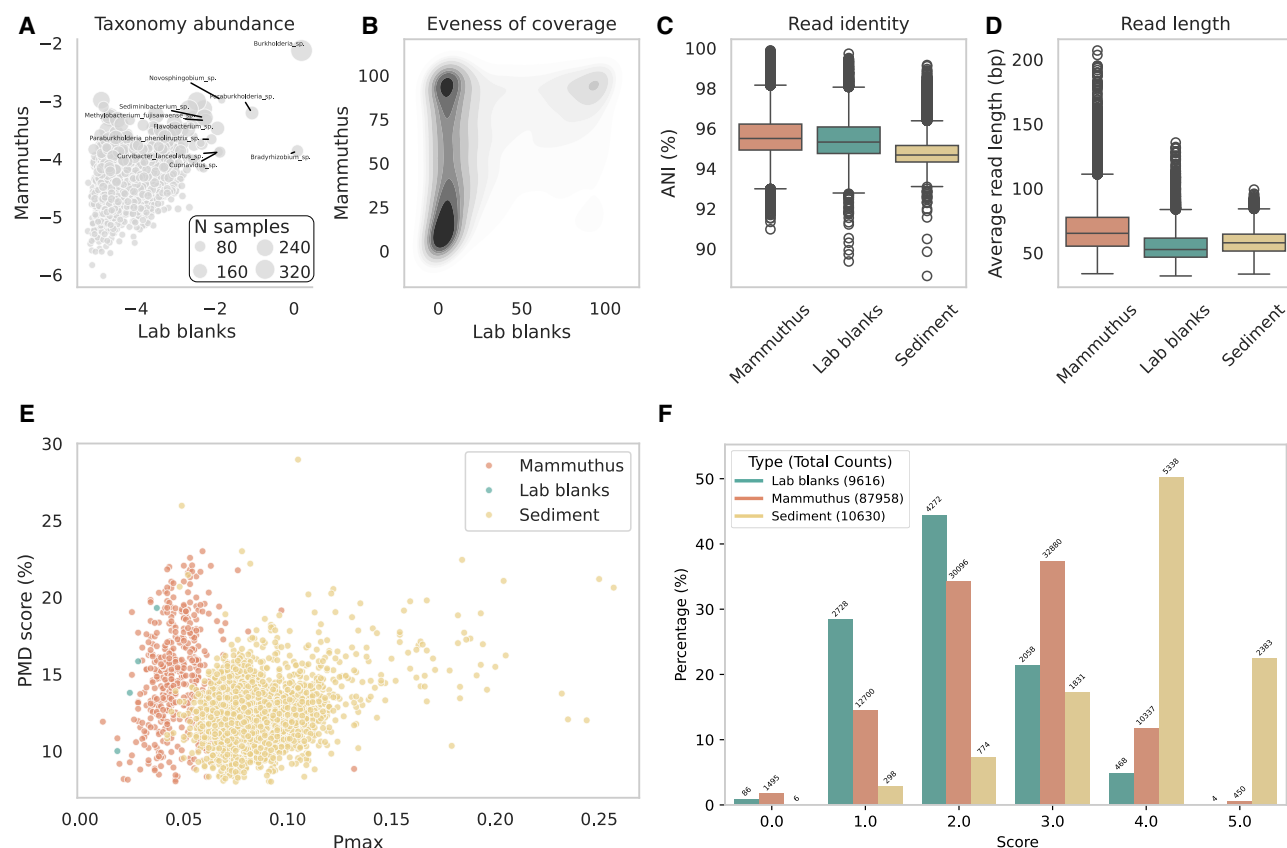


Figure 1. General plots about read characteristics and microbial classification

In all sections, the colors yellow, green, and brown stand for sedimentary, laboratory blanks, and mammoth reads, respectively.

(A) Illustrates the microbial taxonomy abundance of *Kraken2* classified reads among all mammoth samples compared with the same taxa in the laboratory blanks. Each circle represents one microbe species classified from GTDB. Both axes represent the log10 values of the ratio of minimizers to total sequenced reads in the sample.

(B) Illustrates the evenness of coverage across contamination samples, including blanks in comparison to mammoth samples, with darker regions indicating higher density of data points.

(C and D) Represent boxplots showing the distribution of average nucleotide identity (ANI) and mean lengths of reads mapped to reference assemblies, respectively.

(E) Corresponds to the distribution of Pmax values from the *Pydamage*¹⁶ analysis as a function of post-mortem damage (PMD) scores computed with *PMDtools*¹⁸ for each candidate with a given score of 5.

(F) Corresponds to a bar plot distribution of the percentage of microbes classified according to a scoring system from 1 to 5, where 5 represents the highest score for ancient DNA damage authentication. All associated results can be found within [Table S2](#).

inspected the DNA damage plots of all 2,740 microbial candidates with a score of at least four, an evenness of coverage above 50%, and increasing C-to-T and/or CpG-to-TpG damage at read ends. This resulted in a final set of 310 microbial candidates showing consistent DNA damage patterns (out of the total number of 87,958 microbial candidates classified at the start), distributed across 105 different mammoth samples. Out of 310 microbial species, 140 were derived from mammoth samples with reads showing slight residual C-to-T damage despite UDG treatment, while the remaining sample did not exhibit this residual damage ([Table S3](#)). Among these 310 ancient microbial candidates, the average evenness of coverage across the genome was 85% and the mean depth of coverage 0.78X ([Figure 3A](#)). None of these microbes presented a sequence identity to the phylogenetically closest available reference genome

over 98.8% (mean = 95.60%, SD = 2.74%) ([Figure 3B](#)), suggesting that these microbes do not represent previously sequenced species. However, the overall low coverage and short read length (mean, 59 bp; SD, 26 bp) could also partially account for the low sequence identity observed ([Figure 3C](#)). The majority of the ancient microbes were identified from molar ($n = 164$) and tusk samples ($n = 79$) ([Figure 3E](#)). Most of these ancient microbes come from genera that have previously been isolated from environmental samples, including *Gelidibacter*, *Nitrobacter*, and *Sulfuricella*.^{22–24} This suggests that the majority of the microbes are environmental species that colonized the mammoth remains post-mortem ([Figures 3F and 3G](#); [Table S4](#)). We also detected the presence of *Clostridium* and *Acinetobacter* genera in our samples, typical post-mortem bacteria involved in the decomposition process.^{25,26}

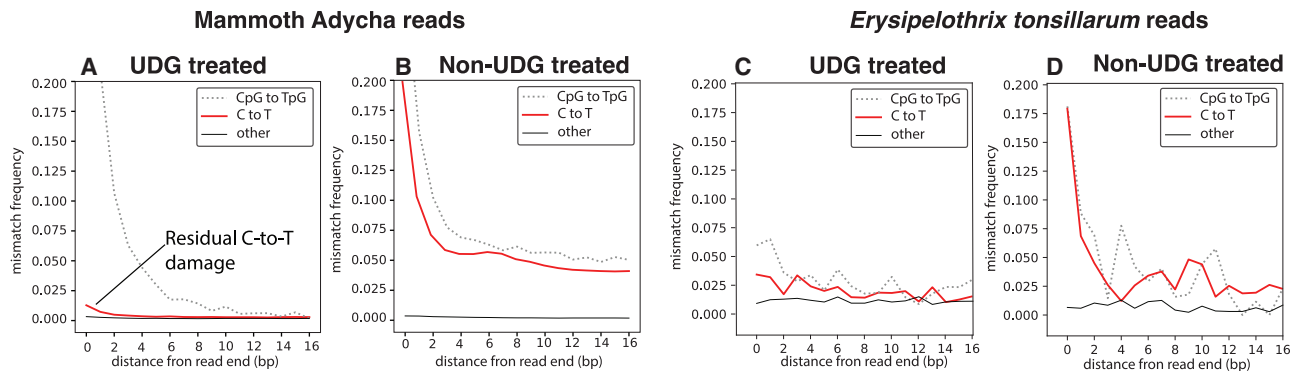


Figure 2. Damage plot distribution of P033 Adycha reads mapped to the Asian elephant and the *Erysipelothrix tonsillarum* assembly

The left side of the figure shows the damage plot from the Adycha mammoth host reads that have been mapped against the Asian elephant genome from (A) UDG-treated library and (B) non-UDG-treated DNA library. The right side of the figure shows the damage plot from Adycha reads that have been mapped against the *Erysipelothrix tonsillarum* assembly (RefSeq: GCF_000373785.1) from (C) UDG-treated library and (D) non-UDG-treated DNA library. Both sequencing types were done on the same mammoth sample (P033) but were processed independently. For the phylogenetic tree inferences, P033 *E. tonsillarum* from non-UDG-treated reads, the reads were trimmed at 10 bp of the edges (*BamUtil* trimBam v1.0.15 -L 10 -R 10 -clip²¹).

See also Figure S4.

Host-associated bacteria

Using the microbe atlas project database²⁷ and literature research, we identified six microbial genera in our filtered data that are known to closely interact with animals (Figure 3F; Table S4). To obtain insights into the evolutionary relationships of these genera to their modern relatives, we systematically inferred genetic phylogenies using all available modern microbial reference genomes from each of these bacterial genera and included all mammoth samples for which these microbes had at least 50% evenness of coverage. While most bacterial candidates contained a high number of sites suitable for phylogenetic analysis, a few, such as *Streptococcus mutans* M25 and *Basfia* strains M40 and M13, had fewer than 10,000 informative sites (Table S11). Given the potential impact of limited coverage on phylogenetic reconstructions, we assessed whether this affected our phylogenetic conclusions by conducting multiple independent phylogeny tests, including building the trees with one mammoth microbe sample at a time, the application of stringent filters on the covered sites, and the use of different phylogenetic placement algorithms (see details in STAR Methods, Data S2; Tables S5 and S6). This allowed us to identify distinct microbial species belonging to the genera *Actinobacillus*/*Basfia*, *Erysipelothrix*, *Odoribacter*, *Pasteurella*, and *Streptococcus* that are commonly associated with animals. All of them, except *Odoribacter*, provided clear phylogenetic evidence of an animal microbial origin, with mammoth bacteria forming monophyletic clades (Figure 4). In the case of *Odoribacter*, we could not rule out post-mortem contamination, as the Arctic sediment reads and other sedimentary bacteria branched within the clade formed by 15 woolly and steppe mammoth *Odoribacter* sequences (Data S3). Furthermore, no clear structure could be found among the mammoth sequences, both in terms of expected relationships among mammoth species or geological age, which means these sequences probably did not interact with mammoths but rather occupied an ecological niche in the sediment. For each of the other host-associated bacteria

genera, we observed their occurrence within mammoth samples across a wide time range and different geographic regions, suggestive of a long-term evolutionary host-association. The microbial genome (*Erysipelothrix*) was isolated from the 1.1-million-year-old steppe mammoth data, representing, to our knowledge, the oldest host-associated microbial genome recovered to date.

Two distinct clades of *Pasteurella* were recovered in 11 woolly mammoths from different geographic locations (Wrangel, Yukon, Ayon Island, Muus-Khaya, and the Taimyr Peninsula), spanning from the Late Pleistocene to 4,607 years ago (Figure 4A). One of the clades is phylogenetically close to an *Actinobacillus* bacterium previously isolated from pig (*Sus domesticus*) feces, which is also related to the bacterium *Basfia succiniciproducens* isolated from domestic cattle (*Bos taurus*) and known to play a function in the production of succinic acid. Succinic acid plays a crucial role in anaerobic fermentation, which is particularly relevant for herbivorous mammals like mammoths that relied on microbial fermentation to break down plant material in their digestive systems.³¹ However, as the mammoth clade is genetically distant from that of *Basfia succiniciproducens*, the exact biological function of this bacterium in mammoths may have been different. Since this bacterial group is found within the oral cavity of their hosts,³² and given that we detected them exclusively within mammoth molar samples (and none in the other 280 tissues from non-molar origin), these bacteria were likely oral commensals that co-existed with woolly mammoths in the Holocene and at least during the Late Pleistocene (Figure 4A). Notably, the most basal *Actinobacillus*-like strain found in mammoths also corresponds to the oldest mammoth specimen of this clade (FK012), dated to the Late Pleistocene. Furthermore, all ancient bacteria in this clade showed a higher sequence similarity to a modern outgroup bacterium (*Basfia succiniciproducens*) than the modern genomes did (Figure 6B), aligning with the expectation that ancient bacteria have accumulated fewer derived mutations to an outgroup than their modern counterparts as less evolutionary time has passed in ancient samples (Figure 6A).

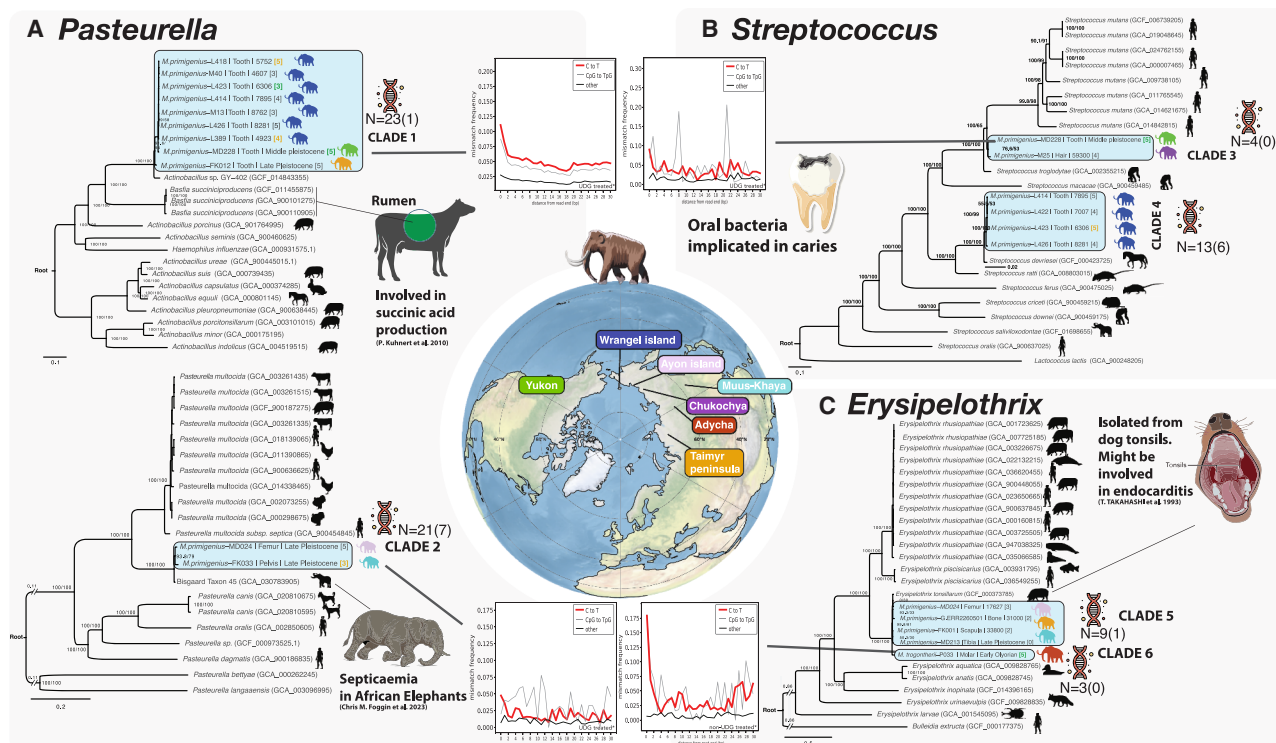


Figure 4. Phylogenetic inference of the four mammoth-associated bacterial genera

Maximum likelihood phylogenetic trees, including *Pasteurella* (A), *Streptococcus* (B), and *Erysipelothrix* (C), were inferred with IQ-TREE²⁸ from a nucleotide genome alignment performed with PanAcoTA v.1.4.1.²⁹ Confidence scores (aLRT%/ultra-bootstrap support%) are shown at each node. A branch-length scale is displayed at the bottom of each phylogeny. The trees were rooted using species of a related outgroup genus. Icons next to each node refer to the host species from which the assembly was isolated (when known). The colors in the mammoth icons refer to the regions where the samples were found, and the exact localizations can be found in the circular map in the center (green: Yukon-Canada, blue: Wrangel Island-Russia; purple: Chukochya-Russia, pink: Ayon Island-Russia, cyan: Muus-Kahya-Russia, red: Adycha-Russia, and orange: Taimyr Peninsula-Russia). The mammoth tip labels show the following sample information (identification name | tissue source | estimated age of the sample [ancient authentication score where black means no damage observed, orange means putative damage and green means confident C-to-T damage observed]). For each modern bacterium, the assembly GenBank ID is shown between brackets in the tip label name. Each illustration adjacent to the phylogenies depicts a documented interaction effect of a bacterium from the same genus in other animal hosts. Next to each mammoth clade the DNA molecule illustrates the number of virulent gene matches with the virulence factor database, and between brackets the number of confirmed virulent genes using VirulentPred v.2.0.³⁰ All reads mapped to reference genome information including damage, read length, percentage identity, genome coverage can be found within Data S4. All reads' average nucleotide identity (ANI) and read percentage identity of the reads selected to build the phylogenies can be found in Data S4. The damage pattern plot without UD treatment of the *Erysipelothrix* sequenced from the *M. trogontherii* can be found in Figure 2. All newick tree files and number of sites covered for each mammoth species can be found in Table S5A. The pairwise ANI between the mammoth ancient strain and their closest sequences within the phylogenies can be found within Table S10. The total number of sites shared by mammoth bacteria within the alignments can be found within Table S11.

See also Figures S5 and S6.

S. mutans strain (RefSeq: GCF_006739205.1), looking for genes with no match to any other *Streptococcus* genome. This way, we identified 16 genes that are unique to the human *S. mutans* species (see Table S7), and none of these 16 genes were detected in the mammoth data. The probability of not sequencing any of these 16 core genes by chance was calculated to be below 0.0188 (see STAR Methods and Table S8 for more details), thus providing further support that the *Streptococcus* isolated from mammoth represents a unique species distinct from human-associated *S. mutans*. However, due to low coverage, the bootstrap support for the phylogenetic placement of these two *Streptococcus* mammoth strains remains low, and the various tests we conducted to determine their exact phylogenetic placement were inconclusive (see STAR Methods and Data S2, sec-

tion B). These sequences might have been related to *S. dentiloxodontae*, which is a *Streptococcus* isolated from elephants and might be closely related to the *S. mutans*/*S. troglodytae*.³⁵ However, we conducted a blastn against the *gyrB* gene in *S. dentiloxodontae* and found a highest identity to the *S. mutans* strain (E-value = 5e-94, percent identity = 94.67%), while the *S. dentiloxodontae* sequence was more distant (E-value = 3e-32, percent identity = 75.11%). Therefore, while we can confirm that the reads do not belong to *S. mutans* or *S. dentiloxodontae*, we cannot rule out contamination from other unknown human sources. This is suggested by the placement of the mammoth clade between human and chimpanzee *Streptococcus*, as well as by the hair tissue origin of M25, which is unlikely to preserve bacterial remains.

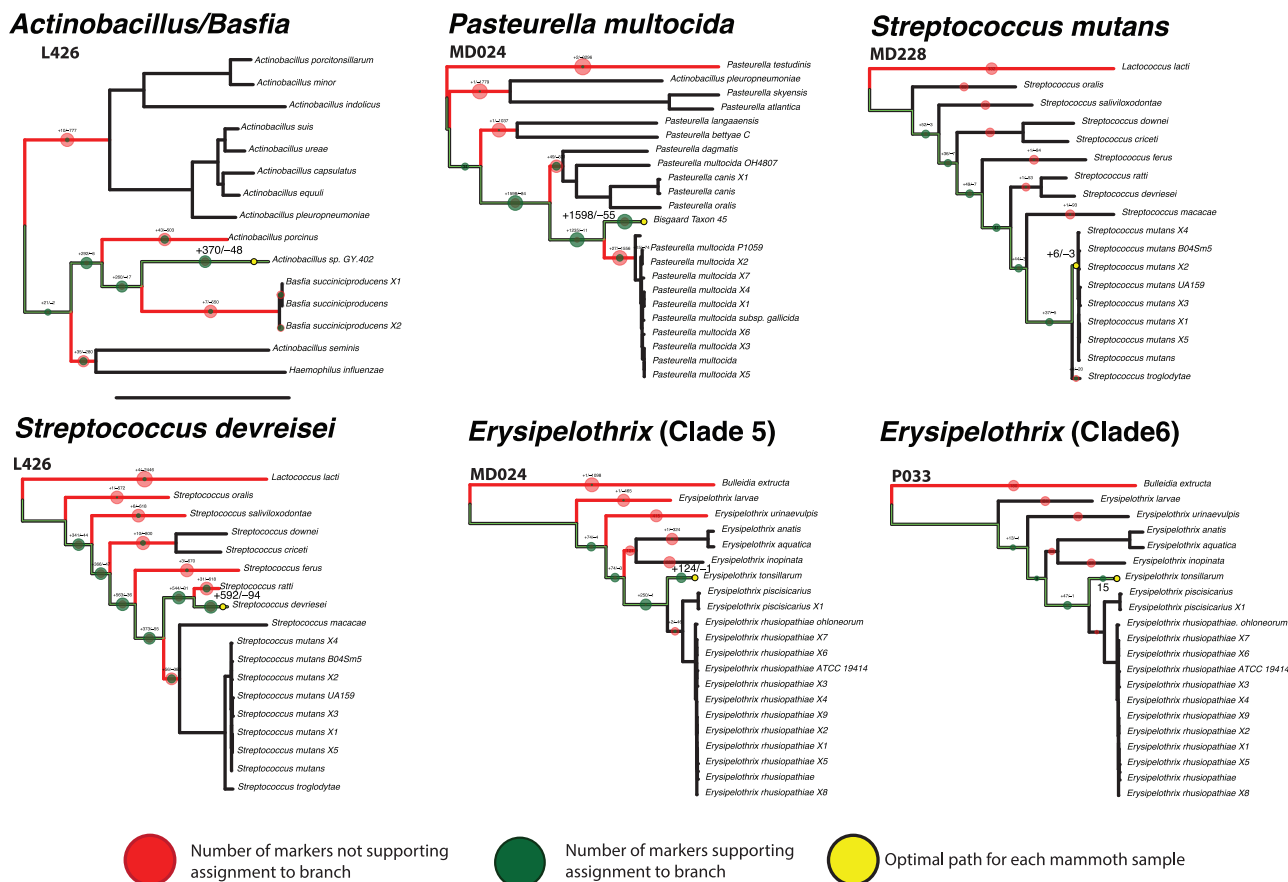


Figure 5. pathPhynder phylogenetic placement of selected candidates for each ancient mammoth bacterium

Phylogenetic trees were inferred using gene alignment output from PanACoTA and reads were mapped using bowtie2 (–very sensitive, –q30). The clade number listed next to each bacterial genus corresponds to its phylogenetic clade number shown in Figure 4. The red and green circles correspond to the number of markers that do not support and do support the assignment to the branch, respectively. The yellow circle corresponds to the optimal assignment path for each sample. All the individual mammoth microbial pathPhynder phylogenetic placement analysis can be found within Data S2.

In addition, four other *Streptococcus*-like sequences isolated from Wrangel Island woolly mammoth teeth formed a monophyletic clade and robust to all our phylogenetic tests (see STAR Methods and Data S2, section C). These sequences are related to *Streptococcus devriesei*, a microbe causing dental caries in horses.³⁶ All of these *Streptococcus devriesei*-like sequences exhibited less dissimilarity to an outgroup bacterium (*Streptococcus rattus*) compared with the modern *S. devriesei* genomes further supporting their ancient origin (Figure 6B). These four mammoth-derived bacterial species did not show a close relationship to a *Streptococcus* species previously found in African elephants (*S. salivoxodontae*) (Figure 4B). Additionally, 16S sequences available for three other *Streptococcus* species found in African elephants, (*S. oriloxodontae*, *S. loxodontisalivarius*, and *S. dentiloxodontae*)^{35,37,38} did not show a close relationship to *S. devriesei* based on 16S rRNA phylogenetic analysis. This indicates that these newly discovered mammoth-associated *Streptococcus* species are not closely related to those found in modern elephants.

Finally, we identified a clade of *Erysipelothrix* isolated from four woolly mammoth bones and a molar of a steppe mammoth (Ady-

cha) (Figure 4C). Among these, four strains were found to be related to the *Erysipelothrix tonsillarum* previously isolated from pig tonsil and dogs (*Canis familiaris*).^{39,40} The most deeply divergent *E. tonsillarum*-related lineage was found in the ~1.1-million-year-old steppe mammoth (Figure 4C). The *Erysipelothrix* found within the steppe mammoth was branching as an outgroup compared with the other woolly mammoth *Erysipelothrix*-like bacteria related to *E. tonsillarum*, suggesting that two distinct bacteria were present between the two mammoth species. Furthermore, the woolly mammoth bacterium exhibited an average nucleotide identity of 99.2% with *E. tonsillarum*, while the 1.1-million-year-old *Erysipelothrix* was more distantly related (average nucleotide identity = 98.8%) (Table S10). The substantial divergence of over 1 million years between the *Erysipelothrix tonsillarum* genome obtained from the Adycha sample and the genomes of *E. tonsillarum* from other mammoth samples suggests considerable evolutionary change has occurred between these strains. To confirm that the 1.1-million-year-old genome is indeed a distinct species, we assessed the sequence divergence by comparing the gene alignments between the different *E. tonsillarum* genomes. On average, 1.24% of the sites differ

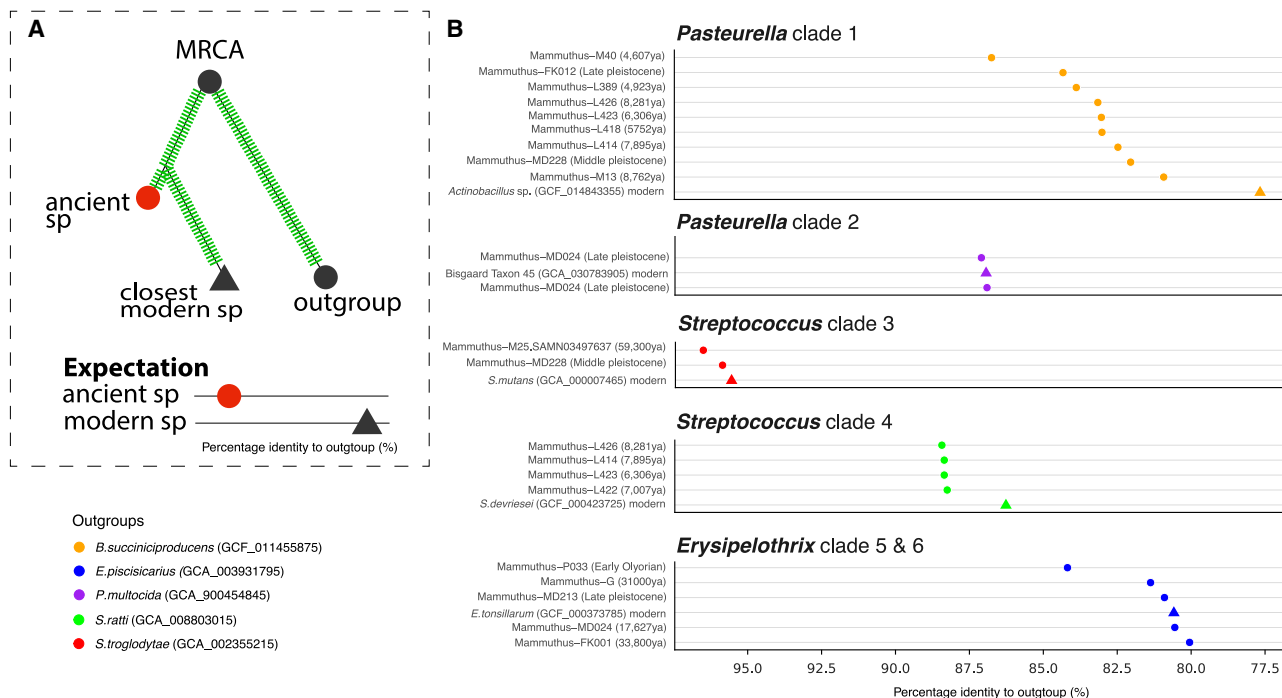


Figure 6. Comparative analysis of sequence dissimilarity between ancient microbial sequences and modern bacteria relative to their closest outgroups

(A and B) Schematic illustrating the hypothesis: putative ancient microbial sequences are expected to show lower dissimilarity to the closest outgroup compared with modern bacterial sequences, due to shorter evolutionary divergence. Green bars along the branches illustrate mutations.

(B) Results showing the percentage of identity between ancient and modern microbial sequences compared with modern outgroups across the clades shown in Figure 4. A color is assigned to each different outgroup species where circles and triangles represent ancient and modern microbial species respectively. The estimated age of each mammoth sample is indicated in brackets.

between the 1.1-million-year-old genome and the genomes from the other mammoth samples (Figure S3). To further contextualize this divergence, we measured the sequence identity for 25 core genes across a large phylogeny based on 10,575 different microbial species.⁴¹ Despite variability in the rate of evolution across different genes, we found that a 1.24% sequence divergence between the 1.1-million-year-old genome and the Late Pleistocene genomes falls within the typical range observed for divergences between different microbial species (Figure S3B). This high divergence to the modern reference also supports the ancient origin of the 1.1-million-year-old *E. tonsillarum* genome. In addition to this analysis and the high damage pattern observed in the steppe mammoth *Erysipelothrix* bacteria (Figure 2D), we also find that the *Erysipelothrix*-like bacteria found in the mammoths within clade 6 accumulated fewer derived mutations compared with the modern *E. tonsillarum* strains when compared to the same outgroup (*E. piscisarius*) (Figures 6A and 6B). In particular, the 1.1-million-year-old genome contained much fewer derived alleles, further supporting the ancient origin of this bacterium (Figure 6B).

Virulence gene content

To investigate the pathogenic potential of the candidate host-associated ancient microbial clades, we pooled all reads within each clade and reconstructed consensus genomes using a reference-based approach (see STAR Methods, virulent gene

identification for more details). Specifically, all merged reads were mapped to the closest modern reference genome identified based on the phylogeny in Figure 4. This method inherently relies on the presence of homologous genes in modern reference genomes, meaning that truly unique genetic elements of the ancient strain may not be fully represented. We recovered substantial portions of the genomes for *E. tonsillarum* clade 5 (88%), Bisgaard taxon 45 clade 2 (50%), *S. devriesei* clade 4 (28.5%), *Actinobacillus* clade 1 (11.5%), *E. tonsillarum* clade 6 (9%), and *S. mutans* clade 3 (1.8%).

Screening these bacterial assemblies identified 81 putative virulence genes, 15 of which were at high confidence due to having near-complete recovered target protein sequences (Table S9; Figure S4). The *Pasteurella* strain clade 2 presented the most virulent genes detected with 21 virulence gene matches, among which seven present a coverage over 90%. Among them we found a lipopolysaccharide heptosyltransferase II (96.5% coverage and 2.089e–80 E-value) (Table S9), which is known as a critical virulence determinant in *Pasteurella multocida* and a major antigen responsible for host protective immunity,⁴² as well as a gene involved in lipopolysaccharide inner-core biosynthesis (*HldE*, coverage = 99.7% and 4.8e–17 E-value) (Table S9), which is an important factor involved in the pathogenesis and virulence for many bacteria including *Actinobacillus pleuropneumoniae*.^{43,44} In *Streptococcus devriesei* clade 4, we found 13

putative virulence genes. Among them, six presented a coverage over 90%. One of these is a dTDP-4-dehydrorhamnose reductases (coverage = 92.5%, 7.758e–107 E-value) (Table S9), which is known to be implicated into the viability or virulence of bacteria including *Streptococcus*.⁴⁵ Another, rhamnose-glucose polysaccharide assembly protein RgpF (coverage = 100%, 1.146e–186 E-value) (Table S9), has been identified as critical in the maintenance of overall stress tolerance and virulence traits in *Streptococcus mutans*.^{45,46} In *E. tonsillarum* clade 5, we found nine matches with virulent genes. Among them, one presented a coverage over 90% with strong homology (coverage = 0.99%, E-value = 0) to the *RspB* protein (Table S9). This protein has been previously identified in *E. rhusiopathiae* and is involved in biofilm initiation by binding to both abiotic and biotic surfaces. Biofilm formation enhances bacterial colonization and reduces antibiotic effectiveness and host defenses, making it a crucial virulence factor in many bacteria.⁴⁷ Although we identified genes commonly associated with infectivity, caution is needed as the mere presence of virulence factors does not necessarily confirm the pathogenicity of the microbes as some genes can also be found in commensal or opportunistic environmental strains.^{48,49} Moreover, some bacteria that carry genes associated with increased pathogenicity may still behave as commensals within the host species and do not always cause disease. For instance, *Streptococcus mutans* and *Erysipelothrix tonsillarum* are typically found as commensal bacteria in humans and dogs, respectively, but can occasionally exhibit pathogenic effects under certain conditions in the same or other hosts.^{39,50}

DISCUSSION

We obtained microbial genomic data from mammoth remains spanning a broad geographical range and a temporal scale from the Holocene (~4,000 years ago) to the Early Pleistocene, over a million years ago. Most of the bacteria identified in this study were also found in ancient Arctic sediment samples, suggesting that post-mortem colonization of the mammoth tissues was the main source of the detected species. However, through extensive bioinformatic filtering, we identified several instances of host-associated ancient bacteria, which were primarily obtained from mammoth teeth.

The most common host-associated clade we detected was related to *Actinobacillus/Basfia* species previously isolated from pig fecal samples. Since related bacteria are commonly found in the oral cavities of their hosts,³² and all *Actinobacillus/Basfia* data in our study were obtained from mammoth teeth, we propose that these microbes were part of the mammoth oral microbiome, likely existing as commensals. This can be further explained by the observation that teeth, as part of the skeleton, are often colonized by microbes due to their high vascularization and porous structure.⁵¹ In future studies, targeting dental calculus directly from mammoths' teeth, when present, could yield higher amounts of ancient microbial sequences, as dental calculus is known to be a rich source of ancient microbial DNA.⁵² The phylogenetic analysis revealed a well-supported monophyletic clade of the bacteria isolated from Wrangel Island mammoths with ages ranging from 4,607 to 8,763 years ago and a distinct, divergent clade in woolly mammoths from the Late and

Middle Pleistocene. This clustering pattern, in which older strains form basal branches within the clade, along with their high genetic divergence from contemporary strains (95.16% identity, SD = 0.44), suggests that these bacterial lineages co-existed with mammoths throughout their evolutionary history.

We also identified ancient bacteria closely related to *Erysipelothrix tonsillarum*, previously isolated from the tonsils of pigs and dogs, where it has been implicated in endocarditis, a condition that typically arises when bacteria enter the bloodstream and colonize internal tissues.³⁹ Notably, despite the majority of our mammoth dataset originating from teeth, all woolly mammoth *Erysipelothrix*-like candidates within clade 5 were exclusively found in bone-tissue samples (tibia, scapula, and femur). This suggests that *E. tonsillarum* may occasionally reach bone tissue through the bloodstream, either as part of a systemic infection or as a commensal microbe that translocates to deeper tissues, potentially leading to osteoarticular infections as observed in *E. rhusiopathiae*.⁵³ Alternatively, it may have been a skin-associated bacterium that infiltrated the skeletal remains post-mortem. Additionally, we identified a bacterium related to *E. tonsillarum* clade 6 from a 1.1-million-year-old steppe mammoth, representing the oldest zoonotic bacterium sequenced to date. However, all the mammoth *Erysipelothrix* bacteria did not form a supported single monophyletic clade, making it difficult to confirm the bacterium's persistence across speciation in woolly mammoths and the steppe mammoths. Interestingly, the post-mortem C-to-T damage profile of this 1.1-million-year-old strain closely mirrored that of the steppe mammoth genome, indicating similar patterns of DNA degradation and suggesting that these bacteria were indeed contemporaneous with their host. While some *Erysipelothrix* species such as *E. rhusiopathiae* can persist in soil, they require a vertebrate host to complete their life cycle.^{39,54,55} The presence of *Erysipelothrix* within a clade encompassing mammoths from various geological periods and locations therefore suggests a long-term association with mammoths or a shared habitat through time.

Bisgaard taxon 45 was recently identified as the cause of multiple African elephant deaths in Zimbabwe.³³ In our results, we identified a bacterium strain related to this Bisgaard taxon 45 in two woolly mammoth samples, excavated in Russia and dating back to the Late Pleistocene period. Bisgaard 45 typically resides in the respiratory tracts and oral cavities of various animals, including domestic species.^{33,56} It can either exist harmlessly as a commensal or, conversely, become a dangerous pathogen causing severe conditions like septicemia.⁵⁷ While Bisgaard taxon 45 is currently the only elephant strain for which a genome is available, hemorrhagic septicemia has also been reported in Asian elephants (*Elephas maximus*) across Sri Lanka, India, and Thailand.^{57–59} Given the potential impact of Bisgaard taxon 45 on elephants, which can cause septicemia and damage to multiple organs, including the lungs, liver, and stomach,^{33,60} and the relative high number of reads detected in the two mammoth samples, one might hypothesize that these mammoth individuals may have experienced a septicemic stage. However, similar to most other *Pasteurella* bacteria, Bisgaard taxon 45 is likely an opportunistic pathogen that persists as a commensal in its hosts.⁶¹ Therefore, its exact role in mammoth health will require further validation.

The detection of host-associated bacteria in mammoths demonstrates that these ancient samples can provide insights into the microbial communities that co-existed together with their host. Such results have previously also been obtained for ancient horse, chicken, or rodent samples, for example.^{62–65} However, several factors constrain our ability to draw detailed conclusions about bacterial prevalence in ancient mammoth populations. These include the inherently low DNA quantities obtained for most of the microbes, post-mortem colonization of the samples by environmental bacteria requiring conservative filtering methods, likely also removing true signals, and the limited availability of appropriate reference genomes in existing microbial databases. Despite these challenges, our findings form a basis for further research toward a deeper understanding of the microbiome and their impact on health and diseases in Pleistocene megafauna.

Limitations of the study

A limitation of this study was the nature of the mammoth sequencing data, which was originally obtained for studying the host genome and therefore the majority of samples were treated with UDG. While UDG treatment enhances sequence reliability, it also removes post-mortem damage patterns that aid in the characterization of ancient microbial DNA. As a result, our ability to confirm the ancient origin of most bacterial candidates remained restricted, and most lineages were excluded for further analysis in this study. A more effective approach would involve half-UDG treatment, which removes most molecular damage, while preserving C-to-T transitions at the read termini. This retained damage could then be selectively removed bioinformatically, allowing for precise characterization of the host genome while maintaining ancient DNA damage patterns for the validation of ancient microbes.⁶⁶ Another limitation of our study was the use of proxy samples from nearby ancient sediments for post-mortem contaminant filtering rather than having sediments from the immediate environment surrounding the same mammoth remains. To enhance future sampling strategies, where possible, collecting not only the primary sample but also contextual environmental samples, such as soil from the excavation site, could serve as a more informative proxy for assessing bacterial contamination. Finally, for the majority of microbial genomes in this study only part of the genome was recovered. Since most of these represent newly sequenced species, it remains unclear what the exact role of these was in relation to their mammoth host. In future work, additional sequence data could be obtained by designing capture baits for each of the identified species, which could further improve phylogenomic resolution, as well as the detection of functional genes including those important for virulence of the species, thus allowing for a broader characterization of the microbiome function profile. This would also make it possible to detect gene-selection signals allowing for studies on microbe-host co-evolution through time.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Benjamin Guinet (benjamin.guinet95@gmail.com).

Materials availability

This study did not generate any new unique reagents. Access to mammoth DNA extracts is available upon request.

Data and code availability

- All data needed to evaluate the conclusions in the paper are present in the paper and/or the [supplemental materials](#).
- All merged FASTQ reads that did not align against the Asian elephant reference genome (GCF_024166365.1) can be found in the European Nucleotide Archive (ENA): PRJEB78615.
- All scripts used in this paper can be found at https://github.com/BenjaminGuinet/Mammoth_Metagenomics.

ACKNOWLEDGMENTS

We would like to thank Zoé Pochon for helpful discussions and Alexandre Gilar-det, Isabelle Feinauer, and Thijessen Naidoo for providing laboratory blanks. We thank Julia Höglund for making the mammoth illustrations in [Figure 4](#). We thank the Vuntut Gwitchin and Tr'ondëk Hwëch'in for their stewardship of the lands where Yukon fossils were recovered and for their support of our palaeontology research. Thank you to the Canada Polar Continental Shelf Program for its support of research to collect fossils in the Old Crow region. Thank you to the Klondike gold mining community for their support of fossil research at Yukon mines. T.v.d.V. and B.G. acknowledge support from the SciLifeLab and Wallenberg Data Driven Life Science Program (KAW 2020.0239). N.O. is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. L.D. acknowledges support from the Swedish Research Council (2017-04647 and 2021-00625) and the European Union (ERC, PrimiGenomes, 101054984). P.P. received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 892446. J.C.C.-D. acknowledges funding from the European Union's Horizon Europe programme under the Marie Skłodowska-Curie Actions Postdoctoral Fellowships (101111414). The authors also acknowledge support from Science for Life Laboratory (SciLifeLab), the National Genomics Infrastructure (NGI) funded by the Swedish Research Council, and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for access to the UPPMAX computational infrastructure.

AUTHOR CONTRIBUTIONS

Conceptualization, B.G. and T.v.d.V.; methodology, B.G. and T.v.d.V.; investigation, B.G. and T.v.d.V.; visualization, B.G. and T.v.d.V.; funding acquisition, T.v.d.V.; project administration, T.v.d.V.; supervision, T.v.d.V.; writing – original draft, B.G.; writing – review & editing, B.G., T.v.d.V., N.O., K.M., M.D., J.C.C.-D., A.A., J.L.A., G.D., F.K., A.C.K., H.M., V.P., A.P., A.T., L.T., G.Z., P.M., L.G., M.R., B.S., A.M.L., S.V., D.D.-d.-M., A.G., P.P., P.N., and L.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [METHOD DETAILS](#)
 - Sample collection and processing
 - Metagenomic screening
 - Contaminating assessment
 - Authentication of ancient DNA
 - Ecological niche analysis of microbes
 - Visualization and plot analysis
 - Phylogenetic inference
 - Phylogenetic robustness analysis
 - Divergence analysis of ancient microbial sequences

- Virulent gene identification
- Microbial species divergence analysis of *Erysipelothrix*
- *S. mutans* gene content characterization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2025.08.003>.

Received: December 6, 2024

Revised: April 26, 2025

Accepted: August 4, 2025

REFERENCES

- Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., et al. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* 25, 1395–1400. <https://doi.org/10.1016/j.cub.2015.04.007>.
- Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A.M., To, T.-H., Kortschak, R.D., et al. (2018). A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci. USA* 115, E2566–E2574. <https://doi.org/10.1073/pnas.1720554115>.
- van der Valk, T., Pečnerová, P., Díez-Del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J.A., Dehasque, M., Sağlıcan, E., et al. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 591, 265–269. <https://doi.org/10.1038/s41586-021-03224-9>.
- Dehasque, M., Morales, H.E., Díez-Del-Molino, D., Pečnerová, P., Chacón-Duque, J.C., Kanellidou, F., Muller, H., Plotnikov, V., Protopopov, A., Tikhonov, A., et al. (2024). Temporal dynamics of woolly mammoth genome erosion prior to extinction. *Cell* 187, 3531–3540. <https://doi.org/10.1016/j.cell.2024.05.033>.
- Díez-Del-Molino, D., Dehasque, M., Chacón-Duque, J.C., Pečnerová, P., Tikhonov, A., Protopopov, A., Plotnikov, V., Kanellidou, F., Nikolskiy, P., Mortensen, P., et al. (2023). Genomics of adaptive evolution in the woolly mammoth. *Curr. Biol.* 33, 1753–1764. <https://doi.org/10.1016/j.cub.2023.03.084>.
- Burkhardt, S., Hentschke, J., Weiler, H., Ehlers, B., Ochs, A., Walter, J., Wittstatt, U., and Göltenboth, R. (1999). Elephant herpes virus—a problem for breeding and housing of elephants. *Berl. Munch. Tierarztl. Wochenschr.* 112, 174–179.
- Ossent, P., Guscetti, F., Metzler, A.E., Lang, E.M., Rübel, A., and Hauser, B. (1990). Acute and fatal herpesvirus infection in a young Asian elephant (*Elephas maximus*). *Vet. Pathol.* 27, 131–133. <https://doi.org/10.1177/030098589002700212>.
- Walsh, M.G., Mor, S.M., and Hossain, S. (2019). The elephant–livestock interface modulates anthrax suitability in India. *Proc. Biol. Sci.* 286, 20190179. <https://doi.org/10.1098/rspb.2019.0179>.
- Richman, L.K., Montali, R.J., Garber, R.L., Kennedy, M.A., Lehnhardt, J., Hildebrandt, T., Schmitt, D., Hardy, D., Alcendor, D.J., and Hayward, G. S. (1999). Novel endotheliotropic herpesviruses fatal for Asian and African elephants. *Science* 283, 1171–1176. <https://doi.org/10.1126/science.283.5405.1171>.
- Spyrou, M.A., Bos, K.I., Herbig, A., and Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* 20, 323–340. <https://doi.org/10.1038/s41576-019-0119-1>.
- Aming, N., and Wilson, D.J. (2020). The past, present and future of ancient bacterial DNA. *Microb. Genom.* 6, mgen000384. <https://doi.org/10.1099/mgen.0.000384>.
- 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. (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506–510. <https://doi.org/10.1038/nature10549>.
- 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. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* 46, 336–344. <https://doi.org/10.1038/ng.2906>.
- Feuerborn, T.R., Palkopoulou, E., van der Valk, T., von Seth, J., Munters, A.R., Pečnerová, P., Dehasque, M., Ureña, I., Ersmark, E., Lagerholm, V. K., et al. (2020). Competitive mapping allows for the identification and exclusion of human DNA contamination in ancient faunal genomic datasets. *BMC Genomics* 21, 844. <https://doi.org/10.1186/s12864-020-07229-y>.
- Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 132. <https://doi.org/10.1186/s13059-016-0997-x>.
- Courtin, J., Perfumo, A., Andreev, A.A., Opel, T., Stoof-Leichsenring, K.R., Edwards, M.E., Murton, J.B., and Herzschiuh, U. (2022). Pleistocene glacial and interglacial ecosystems inferred from ancient DNA analyses of permafrost sediments from Batagay megaslump, East Siberia. *Environ. DNA* 4, 1265–1283. <https://doi.org/10.1002/edn3.336>.
- Rigou, S., Christo-Foroux, E., Santini, S., Goncharov, A., Strauss, J., Grosse, G., Fedorov, A.N., Labadie, K., Abergel, C., and Claverie, J.-M. (2022). Metagenomic survey of the microbiome of ancient Siberian permafrost and modern Kamchatkan cryosols. *Microlife* 3, uqac003. <https://doi.org/10.1093/femsml/uqac003>.
- 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. *Proc. Natl. Acad. Sci. USA* 111, 2229–2234. <https://doi.org/10.1073/pnas.1318934111>.
- Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38, e87. <https://doi.org/10.1093/nar/gkp1163>.
- Wojciechowski, M., Czapinska, H., and Bochtler, M. (2013). CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.Mpel. *Proc. Natl. Acad. Sci. USA* 110, 105–110. <https://doi.org/10.1073/pnas.1207986110>.
- Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 25, 918–925. <https://doi.org/10.1101/gr.176552.114>.
- Bowman, J.P. (2006). The Marine Clade of the Family Flavobacteriaceae: The Genera *Aequorivita*, *Arenibacter*, *Cellulophaga*, *Croceibacter*, *Formosa*, *Gelidibacter*, *Gillisia*, *Maribacter*, *Mesonina*, *Muricauda*, *Polariibacter*, *Psychroflexus*, *Psychroserpens*, *Robiginitalea*, *Salegentibacter*, *Tenacibaculum*, *Ulvibacter*, *Vitellibacter* and *Zobellia*. In *The Prokaryotes* (Springer), pp. 677–694.
- Daims, H., Lückner, S., and Wagner, M. (2016). A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. *Trends Microbiol.* 24, 699–712. <https://doi.org/10.1016/j.tim.2016.05.004>.
- Högfors-Rönholm, E., Lundin, D., Brambilla, D., Christel, S., Lopez-Fernandez, M., Lillhonga, T., Engblom, S., Österholm, P., and Dopson, M. (2022). Gallionella and Sulfuricella populations are dominant during the transition of boreal potential to actual acid sulfate soils. *Commun. Earth Environ.* 3, 304. <https://doi.org/10.1038/s43247-022-00642-z>.
- DeBruyn, J.M., and Hauther, K.A. (2017). Postmortem succession of gut microbial communities in deceased human subjects. *PeerJ* 5, e3437. <https://doi.org/10.7717/peerj.3437>.

26. DeBruyn, J.M., Keenan, S.W., and Taylor, L.S. (2025). From carrion to soil: microbial recycling of animal carcasses. *Trends Microbiol.* 33, 194–207. <https://doi.org/10.1016/j.tim.2024.09.003>.
27. Matias Rodrigues, J.F., Schmidt, T.S.B., Tackmann, J., and von Mering, C. (2017). MAPseq: highly efficient k-mer search with confidence estimates, for rRNA sequence analysis. *Bioinformatics* 33, 3808–3810. <https://doi.org/10.1093/bioinformatics/btx517>.
28. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
29. Perrin, A., and Rocha, E.P.C. (2021). PanACoTA: a modular tool for massive microbial comparative genomics. *NAR Genom. Bioinform.* 3, lqaa106. <https://doi.org/10.1093/nargab/lqaa106>.
30. Sharma, A., Garg, A., Ramana, J., and Gupta, D. (2023). VirulentPred 2.0: An improved method for prediction of virulent proteins in bacterial pathogens. *Protein Sci.* 32, e4808. <https://doi.org/10.1002/pro.4808>.
31. Kuhnert, P., Scholten, E., Haefner, S., Mayor, D., and Frey, J. (2010). *Bastia succiniciproducens* gen. nov., sp. nov., a new member of the family Pasteurellaceae isolated from bovine rumen. *Int. J. Syst. Evol. Microbiol.* 60, 44–50. <https://doi.org/10.1099/ijs.0.011809-0>.
32. Bisgaard, M., Phillips, J.E., and Mannheim, W. (1986). Characterization and identification of bovine and ovine Pasteurellaceae isolated from the oral cavity and rumen of apparently normal cattle and sheep. *Acta Pathol. Microbiol. Immunol. Scand. B* 94, 9–17. <https://doi.org/10.1111/j.1699-0463.1986.tb03014.x>.
33. Foggin, C.M., Rosen, L.E., Henton, M.M., Buys, A., Floyd, T., Turner, A.D., Tarbin, J., Lloyd, A.S., Chaitzevi, C., Ellis, R.J., et al. (2023). *Pasteurella* sp. associated with fatal septicaemia in six African elephants. *Nat. Commun.* 14, 6398. <https://doi.org/10.1038/s41467-023-41987-z>.
34. Tanzer, J.M., Livingston, J., and Thompson, A.M. (2001). The microbiology of primary dental caries in humans. *J. Dent. Educ.* 65, 1028–1037. <https://doi.org/10.1002/j.0022-0337.2001.65.10.tb03446.x>.
35. Shinozaki-Kuwahara, N., Saito, M., Hirasawa, M., Hirasawa, M., and Takada, K. (2016). *Streptococcus dentiloxodontae* sp. nov., isolated from the oral cavity of elephants. *Int. J. Syst. Evol. Microbiol.* 66, 3878–3883. <https://doi.org/10.1099/ijs.0.001280>.
36. Collins, M.D., Lundström, T., Welinder-Olsson, C., Hansson, I., Wattle, O., Hudson, R.A., and Falsen, E. (2004). *Streptococcus devriesei* sp. nov., from equine teeth. *Syst. Appl. Microbiol.* 27, 146–150. <https://doi.org/10.1078/072320204322881754>.
37. Saito, M., Shinozaki-Kuwahara, N., Hirasawa, M., and Takada, K. (2014). *Streptococcus loxodontisalivarius* sp. nov. and *Streptococcus saliviloxodontae* sp. nov., isolated from oral cavities of elephants. *Int. J. Syst. Evol. Microbiol.* 64, 3288–3292. <https://doi.org/10.1099/ijs.0.063263-0>.
38. Shinozaki-Kuwahara, N., Saito, M., Hirasawa, M., and Takada, K. (2014). *Streptococcus oriloxodontae* sp. nov., isolated from the oral cavities of elephants. *Int. J. Syst. Evol. Microbiol.* 64, 3755–3759. <https://doi.org/10.1099/ijs.0.064048-0>.
39. Takahashi, T., Fujisawa, T., Benno, Y., Tamura, Y., Sawada, T., Suzuki, S., Muramatsu, M., and Mitsuoka, T. (1987). *Erysipelothrix tonsillarum* sp. nov. Isolated from Tonsils of Apparently Healthy Pigs. *Int. J. Syst. Bacteriol.* 37, 166–168. <https://doi.org/10.1099/00207713-37-2-166>.
40. Takahashi, T., Tamura, Y., Yoshimura, H., Nagamine, N., Kijima, M., Nakamura, M., and Devriese, L.A. (1993). *Erysipelothrix tonsillarum* isolated from dogs with endocarditis in Belgium. *Res. Vet. Sci.* 54, 264–265. [https://doi.org/10.1016/0034-5288\(93\)90071-m](https://doi.org/10.1016/0034-5288(93)90071-m).
41. Zhu, Q., Mai, U., Pfeiffer, W., Janssen, S., Asnicar, F., Sanders, J.G., Belda-Ferre, P., Al-Ghalith, G.A., Kopylova, E., McDonald, D., et al. (2019). Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* 10, 5477. <https://doi.org/10.1038/s41467-019-13443-4>.
42. Harper, M., Boyce, J.D., Cox, A.D., St Michael, F., Wilkie, I.W., Blackall, P. J., and Adler, B. (2007). *Pasteurella multocida* expresses two Lipopolysaccharide glycoforms simultaneously, but only a single form is required for virulence: Identification of two acceptor-specific heptosyl transferases. *Infect. Immun.* 75, 3885–3893. <https://doi.org/10.1128/IAI.00212-07>.
43. Gaddy, J.A., Tomaras, A.P., and Actis, L.A. (2009). The *Acinetobacter baumannii* 19606 OmpA protein plays a role in biofilm formation on abiotic surfaces and in the interaction of this pathogen with eukaryotic cells. *Infect. Immun.* 77, 3150–3160. <https://doi.org/10.1128/IAI.00096-09>.
44. Provost, M., Harel, J., Labrie, J., Sirois, M., and Jacques, M. (2003). Identification, cloning and characterization of *rfaE* of *Actinobacillus pleuropneumoniae* serotype 1, a gene involved in lipopolysaccharide inner-core biosynthesis. *FEMS Microbiol. Lett.* 223, 7–14. [https://doi.org/10.1016/S0378-1097\(03\)00247-7](https://doi.org/10.1016/S0378-1097(03)00247-7).
45. van der Beek, S.L., Le Breton, Y., Ferenbach, A.T., Chapman, R.N., van Aalten, D.M.F., Navratilova, I., Boons, G.-J., McIver, K.S., van Sorge, N. M., and Dorfmueller, H.C. (2015). GacA is essential for Group A *Streptococcus* and defines a new class of monomeric dTDP-4-dehydrohamnose reductases (RmlD). *Mol. Microbiol.* 98, 946–962. <https://doi.org/10.1111/mmi.13169>.
46. Kovacs, C.J., Faustoferri, R.C., and Quivey, R.G., Jr. (2017). RgpF is required for maintenance of stress tolerance and virulence in *Streptococcus mutans*. *J. Bacteriol.* 199, e00497-17. <https://doi.org/10.1128/JB.00497-17>.
47. Shimoji, Y., Ogawa, Y., Osaki, M., Kabeya, H., Maruyama, S., Mikami, T., and Sekizaki, T. (2003). Adhesive surface proteins of *Erysipelothrix rhusiopathiae* bind to polystyrene, fibronectin, and type I and IV collagens. *J. Bacteriol.* 185, 2739–2748. <https://doi.org/10.1128/JB.185.9.2739-2748.2003>.
48. Snyder, L.A.S., and Saunders, N.J. (2006). The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as “virulence genes.” *BMC Genomics* 7, 128. <https://doi.org/10.1186/1471-2164-7-128>.
49. Søborg, D.A., Hendriksen, N.B., Kilian, M., and Kroer, N. (2013). Widespread occurrence of bacterial human virulence determinants in soil and freshwater environments. *Appl. Environ. Microbiol.* 79, 5488–5497. <https://doi.org/10.1128/AEM.01633-13>.
50. Bedoya-Correa, C.M., Rincón Rodríguez, R.J., and Parada-Sanchez, M.T. (2019). Genomic and phenotypic diversity of *Streptococcus mutans*. *J. Oral Biosci.* 61, 22–31. <https://doi.org/10.1016/j.job.2018.11.001>.
51. Fellows Yates, J.A., Warinner, C., Andrades Valtueña, A., Borry, M., Guellil, M., Herbig, A., Hiß, A., Hübner, A., Kocher, A., Lamnidis, T.C., et al. (2024) Introduction to Ancient Metagenomics (Zenodo). <https://doi.org/10.5281/ZENODO.13784555>.
52. Warinner, C., Speller, C., and Collins, M.J. (2015). A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130376. <https://doi.org/10.1098/rstb.2013.0376>.
53. Maillard, A., Wakim, Y., Itani, O., Ousser, F., Bleibtreu, A., Caumes, E., and Monsel, G. (2021). Osteoarticular infections caused by *Erysipelothrix rhusiopathiae*: Case report and literature review. *Open Forum Infect. Dis.* 8, ofab461. <https://doi.org/10.1093/ofid/ofab461>.
54. Forde, T., Biek, R., Zadoks, R., Workentine, M.L., De Buck, J., Kutz, S., Opriessnig, T., Trewby, H., van der Meer, F., and Orsel, K. (2016). Genomic analysis of the multi-host pathogen *Erysipelothrix rhusiopathiae* reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution. *BMC Genomics* 17, 461. <https://doi.org/10.1186/s12864-016-2643-0>.
55. Chandler, D.S., and Craven, J.A. (1980). Persistence and distribution of *Erysipelothrix rhusiopathiae* and bacterial indicator organisms on land used for disposal of piggery effluent. *J. Appl. Bacteriol.* 48, 367–375. <https://doi.org/10.1111/j.1365-2672.1980.tb01024.x>.

56. Ruegg, P.L., and Erskine, R.J. (2020). Mammary gland health and disorders. In *Large Animal Internal Medicine* (Elsevier), pp. 1118–1150. <https://doi.org/10.1016/B978-0-323-55445-9.00036-7>.
57. Christensen, H., and Bisgaard, M. (2024). *Pasteurella*. In *Molecular Medical Microbiology* (Elsevier), pp. 1637–1656. <https://doi.org/10.1016/B978-0-12-818619-0.00094-0>.
58. Chandranaik, B.M., Shivashankar, B.P., Giridhar, P., and Nagaraju, D.N. (2016). Molecular characterisation and serotyping of *Pasteurella multocida* isolates from Asiatic elephants (*Elephas maximus*). *Eur. J. Wildl. Res.* 62, 681–685. <https://doi.org/10.1007/s10344-016-1043-8>.
59. Harish, B.R., Shivaraj, B.M., Chandranaik, B.M., Venkatesh, M.D., and Renukaprasad, C. (2009). Hemorrhagic Septicemia in Asian Elephants *Elephas maximus* in Karnataka state, India. *J. Threat. Taxa* 1, 194–195. <https://doi.org/10.11609/JoTT.o1778.194-5>.
60. Singh, V.N. (2002) Symptomatic Study of Haemorrhagic Septicaemia in Elephant in Mudumalai Wildlife Sanctuary (Tamil Nadu), pp. 1089–1100.
61. Harper, M., Boyce, J.D., and Adler, B. (2006). *Pasteurella multocida* pathogenesis: 125 years after Pasteur. *FEMS Microbiol. Lett.* 265, 1–10. <https://doi.org/10.1111/j.1574-6968.2006.00442.x>.
62. Fiddaman, S.R., Dimopoulos, E.A., Lebrasseur, O., du Plessis, L., Vrancken, B., Charlton, S., Haruda, A.F., Tabbada, K., Flammer, P.G., Dascalu, S., et al. (2023). Ancient chicken remains reveal the origins of virulence in Marek's disease virus. *Science* 382, 1276–1281. <https://doi.org/10.1126/science.adg2238>.
63. Lebrasseur, O., More, K.D., and Orlando, L. (2024). Equine herpesvirus 4 infected domestic horses associated with Sintashta spoke-wheeled chariots around 4,000 years ago. *Virus Evol.* 10, vead087. <https://doi.org/10.1093/ve/vead087>.
64. Mahmoudi, A., Kryštufek, B., Sludsky, A., Schmid, B.V., DE Almeida, A.M. P., Lei, X., Ramasindrazana, B., Bertherat, E., Yeszhanov, A., Stenseth, N. C., et al. (2021). Plague reservoir species throughout the world. *Integr. Zool.* 16, 820–833. <https://doi.org/10.1111/1749-4877.12511>.
65. Pfrengle, S., Neukamm, J., Guellil, M., Keller, M., Molak, M., Avanzi, C., Kushniarevich, A., Montes, N., Neumann, G.U., Reiter, E., et al. (2021). Mycobacterium leprae diversity and population dynamics in medieval Europe from novel ancient genomes. *BMC Biol.* 19, 220. <https://doi.org/10.1186/s12915-021-01120-2>.
66. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130624. <https://doi.org/10.1098/rstb.2013.0624>.
67. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
68. Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794. <https://doi.org/10.1093/nar/gkab776>.
69. Kalyanamamorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermini, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. <https://doi.org/10.1038/nmeth.4285>.
70. Lynch, V.J., Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., Perry, G.H., Miller, W., and Schuster, S.C. (2015). Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the arctic. *Cell Rep.* 12, 217–228. <https://doi.org/10.1016/j.celrep.2015.06.027>.
71. Durvasula, A., Hoffman, P.J., Kent, T.V., Liu, C., Kono, T.J.Y., Morrell, P.L., and Ross-Ibarra, J. (2016). ANGSD-wrapper: utilities for analyzing next generation sequencing data. *Molecular Ecology Resources* 16, 1449–1454. <https://doi.org/10.1111/1755-0998.12578>.
72. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
73. Meyer, M., Palkopoulou, E., Baleka, S., Stiller, M., Penkman, K.E.H., Alt, K. W., Ishida, Y., Mania, D., Mallick, S., Meijer, T., et al. (2017). Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. *eLife* 6, e25413. <https://doi.org/10.7554/eLife.25413>.
74. van der Valk, T., Dehasque, M., Chacón-Duque, J.C., Oskolkov, N., Vartanyan, S., Heintzman, P.D., Pečnerová, P., Díez-Del-Molino, D., and Dalén, L. (2022). Evolutionary consequences of genomic deletions and insertions in the woolly mammoth genome. *iScience* 25, 104826. <https://doi.org/10.1016/j.isci.2022.104826>.
75. Dehasque, M., Pečnerová, P., Kempe Lagerholm, V., Ersmark, E., Danilov, G.K., Mortensen, P., Vartanyan, S., and Dalén, L. (2022). Development and Optimization of a Silica Column-Based Extraction Protocol for Ancient DNA. *Genes* 13, 687. <https://doi.org/10.3390/genes13040687>.
76. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448>.
77. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>.
78. Pochon, Z., Bergfeldt, N., Kirdök, E., Vicente, M., Naidoo, T., van der Valk, T., Altınışık, N.E., Krzewińska, M., Dalén, L., Götherström, A., et al. (2023). aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biol.* 24, 242. <https://doi.org/10.1186/s13059-023-03083-9>.
79. 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. *Annu. Rev. Genomics Hum. Genet.* 18, 321–356. <https://doi.org/10.1146/annurev-genom-091416-035526>.
80. Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., and Rasmussen, S. (2019). Emergence and Spread of Basal Lineages of *Yersinia pestis* during the Neolithic Decline. *Cell* 176, 295–305. <https://doi.org/10.1016/j.cell.2018.11.005>.
81. O'Leary, N.A., Cox, E., Holmes, J.B., Anderson, W.R., Falk, R., Hem, V., Tsuchiya, M.T.N., Schuler, G.D., Zhang, X., Torcivia, J., et al. (2024). Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci. Data* 11, 732. <https://doi.org/10.1038/s41597-024-03571-y>.
82. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
83. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>.
84. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. <https://doi.org/10.1093/molbev/msx281>.
85. Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2019). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst. Biol.* 68, 365–369. <https://doi.org/10.1093/sysbio/syy054>.
86. Czech, L., Barbera, P., and Stamatakis, A. (2020). Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* 36, 3263–3265. <https://doi.org/10.1093/bioinformatics/btaa070>.
87. Martiniano, R., De Sanctis, B., Hallast, P., and Durbin, R. (2022). Placing Ancient DNA Sequences into Reference Phylogenies. *Mol. Biol. Evol.* 39, msac017. <https://doi.org/10.1093/molbev/msac017>.

88. Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., and Jin, Q. (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. <https://doi.org/10.1093/nar/gki008>.
89. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. <https://doi.org/10.1038/nbt.3988>.
90. Dolenz, S., van der Valk, T., Jin, C., Oppenheimer, J., Sharif, M.B., Orlando, L., Shapiro, B., Dalén, L., and Heintzman, P.D. (2024). Unravelling reference bias in ancient DNA datasets. *Bioinformatics* 40, btae436. <https://doi.org/10.1093/bioinformatics/btae436>.
91. Tan, S.Y., Dutta, A., Jakubovics, N.S., Ang, M.Y., Siow, C.C., Mutha, N.V., Heydari, H., Wee, W.Y., Wong, G.J., and Choo, S.W. (2015). YersiniaBase: a genomic resource and analysis platform for comparative analysis of *Yersinia*. *BMC Bioinformatics* 16, 9. <https://doi.org/10.1186/s12859-014-0422-y>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	E464
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	E469D
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L386
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L387
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L414
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L416
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L422
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L423
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	L456
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	M17
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	M32
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	MD076
<i>Mammuthus primigenius</i>	Dehasque et al. ²³	MD080
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	L164
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	L490
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	M9D
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD002
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD018
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD020
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD024
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD039
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD042
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD045
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD046
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	MD090
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	P004
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	P005
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	P007
<i>Mammuthus primigenius</i>	Díez-del-Molino et al. ²⁴	P010
<i>Mammuthus primigenius</i>	Lynch et al. ⁴⁷	M25.SAMN03497637
<i>Mammuthus primigenius</i>	Lynch et al. ⁴⁷	M4.SAMN03497636
<i>Palaeoloxodon antiquus</i>	Meyer et al. ⁵²	Pantiquus-NEPEC.ERR1753652
<i>Palaeoloxodon antiquus</i>	Meyer et al. ⁵²	Pantiquus-NEU2A.ERR1753653
<i>Mammut americanum</i>	Palkopoulou et al. ⁶⁰	M. americanum_I_ERR2260503
<i>Mammut americanum</i>	Palkopoulou et al. ⁶⁰	Mamericanum-X_ERR2260508
<i>Mammuthus primigenius</i>	Palkopoulou et al. ⁶⁰	G_ERR2260501
<i>Mammuthus primigenius</i>	Palkopoulou et al. ⁶⁰	H_ERR2260502
<i>Mammuthus primigenius</i>	Palkopoulou et al. ⁶⁰	S.ERR2260505
<i>Mammuthus</i> sp.	Palkopoulou et al. ⁶⁰	V.ERR2260507
<i>Mammuthus columbi</i>	Palkopoulou et al. ⁶⁰	U.ERR2260506
<i>Palaeoloxodon antiquus</i>	Palkopoulou et al. ⁶⁰	Pantiquus-N.ERR2260504
<i>Mammuthus primigenius</i>	This study	E467
<i>Mammuthus</i> sp.	This study	oimyakon.ERR852028

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus</i> sp.	This study	bcm001
<i>Mammuthus</i> sp.	This study	bcm002
<i>Mammuthus</i> sp.	This study	bcm003
<i>Mammuthus primigenius</i>	This study	bcm004
<i>Mammuthus</i> sp.	This study	bcm005
<i>Mammuthus</i> sp.	This study	bcm006
<i>Mammuthus</i> sp.	This study	bcm007
<i>Mammuthus</i> sp.	This study	bcm008
<i>Mammuthus</i> sp.	This study	bcm009
<i>Mammuthus</i> sp.	This study	bcm010
<i>Mammuthus</i> sp.	This study	bcm011
<i>Mammuthus primigenius</i>	This study	bcm012
<i>Mammuthus</i> sp.	This study	bcm013
<i>Mammuthus</i> sp.	This study	bcm014
<i>Mammuthus imperator</i>	This study	bcm015b
<i>Mammuthus</i> sp.	This study	bcm016
<i>Mammuthus</i> sp.	This study	bcm017
<i>Mammuthus</i> sp.	This study	bcm018
<i>Mammuthus primigenius</i>	This study	bcm019
<i>Mammuthus trogontherii</i>	This study	CC051
<i>Mammuthus trogontherii</i>	This study	CC052
<i>Mammuthus trogontherii</i>	This study	CC053
<i>Mammuthus trogontherii</i>	This study	CC054
<i>Mammuthus trogontherii</i>	This study	CC055
<i>Mammuthus primigenius</i>	This study	E460
<i>Mammuthus primigenius</i>	This study	FK001
<i>Mammuthus primigenius</i>	This study	FK002
<i>Mammuthus primigenius</i>	This study	FK003
<i>Mammuthus primigenius</i>	This study	FK004
<i>Mammuthus primigenius</i>	This study	FK005
<i>Mammuthus primigenius</i>	This study	FK006
<i>Mammuthus primigenius</i>	This study	FK007
<i>Mammuthus primigenius</i>	This study	FK008
<i>Mammuthus primigenius</i>	This study	FK009
<i>Mammuthus primigenius</i>	This study	FK010
<i>Mammuthus primigenius</i>	This study	FK011
<i>Mammuthus primigenius</i>	This study	FK012
<i>Mammuthus primigenius</i>	This study	FK013
<i>Mammuthus primigenius</i>	This study	FK014
<i>Mammuthus primigenius</i>	This study	FK015
<i>Mammuthus primigenius</i>	This study	FK017
<i>Mammuthus primigenius</i>	This study	FK018
<i>Mammuthus primigenius</i>	This study	FK019
<i>Mammuthus primigenius</i>	This study	FK020
<i>Mammuthus primigenius</i>	This study	FK021
<i>Mammuthus primigenius</i>	This study	FK022
<i>Mammuthus primigenius</i>	This study	FK024
<i>Mammuthus primigenius</i>	This study	FK031

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	FK032
<i>Mammuthus primigenius</i>	This study	FK033
<i>Mammuthus primigenius</i>	This study	FK034
<i>Mammuthus primigenius</i>	This study	FK035
<i>Mammuthus primigenius</i>	This study	HM001
<i>Mammuthus primigenius</i>	This study	HM002
<i>Mammuthus primigenius</i>	This study	HM003
<i>Mammuthus primigenius</i>	This study	HM004
<i>Mammuthus primigenius</i>	This study	HM005
<i>Mammuthus primigenius</i>	This study	HM006
<i>Mammuthus primigenius</i>	This study	HM007
<i>Mammuthus primigenius</i>	This study	HM008
<i>Mammuthus primigenius</i>	This study	HM009
<i>Mammuthus primigenius</i>	This study	HM010
<i>Mammuthus primigenius</i>	This study	HM011
<i>Mammuthus primigenius</i>	This study	HM012
<i>Mammuthus primigenius</i>	This study	HM013
<i>Mammuthus primigenius</i>	This study	HM014
<i>Mammuthus primigenius</i>	This study	HM015
<i>Mammuthus primigenius</i>	This study	HM016
<i>Mammuthus primigenius</i>	This study	HM017
<i>Mammuthus primigenius</i>	This study	HM018
<i>Mammuthus primigenius</i>	This study	L152
<i>Mammuthus primigenius</i>	This study	L155
<i>Mammuthus primigenius</i>	This study	L156
<i>Mammuthus primigenius</i>	This study	L157
<i>Mammuthus primigenius</i>	This study	L158
<i>Mammuthus primigenius</i>	This study	L159B
<i>Mammuthus primigenius</i>	This study	L161
<i>Mammuthus primigenius</i>	This study	L166
<i>Mammuthus primigenius</i>	This study	L268B
<i>Mammuthus primigenius</i>	This study	L269
<i>Mammuthus primigenius</i>	This study	L270
<i>Mammuthus primigenius</i>	This study	L273
<i>Mammuthus primigenius</i>	This study	L278
<i>Mammuthus primigenius</i>	This study	L281
<i>Mammuthus primigenius</i>	This study	L376
<i>Mammuthus primigenius</i>	This study	L377
<i>Mammuthus primigenius</i>	This study	L378
<i>Mammuthus primigenius</i>	This study	L379
<i>Mammuthus primigenius</i>	This study	L380D
<i>Mammuthus primigenius</i>	This study	L381
<i>Mammuthus primigenius</i>	This study	L382
<i>Mammuthus primigenius</i>	This study	L383
<i>Mammuthus primigenius</i>	This study	L384
<i>Mammuthus primigenius</i>	This study	L385
<i>Mammuthus primigenius</i>	This study	L389
<i>Mammuthus primigenius</i>	This study	L390

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	L394
<i>Mammuthus primigenius</i>	This study	L399
<i>Mammuthus primigenius</i>	This study	L400
<i>Mammuthus primigenius</i>	This study	L401
<i>Mammuthus primigenius</i>	This study	L402
<i>Mammuthus primigenius</i>	This study	L403
<i>Mammuthus primigenius</i>	This study	L405
<i>Mammuthus primigenius</i>	This study	L406
<i>Mammuthus primigenius</i>	This study	L407
<i>Mammuthus primigenius</i>	This study	L408
<i>Mammuthus primigenius</i>	This study	L410
<i>Mammuthus primigenius</i>	This study	L411
<i>Mammuthus primigenius</i>	This study	L412
<i>Mammuthus primigenius</i>	This study	L413
<i>Mammuthus primigenius</i>	This study	L417
<i>Mammuthus primigenius</i>	This study	L418
<i>Mammuthus primigenius</i>	This study	L419
<i>Mammuthus primigenius</i>	This study	L420
<i>Mammuthus primigenius</i>	This study	L421
<i>Mammuthus primigenius</i>	This study	L424
<i>Mammuthus primigenius</i>	This study	L425
<i>Mammuthus primigenius</i>	This study	L426
<i>Mammuthus primigenius</i>	This study	L427
<i>Mammuthus primigenius</i>	This study	L428
<i>Mammuthus primigenius</i>	This study	L429
<i>Mammuthus primigenius</i>	This study	L430
<i>Mammuthus primigenius</i>	This study	L431
<i>Mammuthus primigenius</i>	This study	L432
<i>Mammuthus primigenius</i>	This study	L433
<i>Mammuthus primigenius</i>	This study	L434
<i>Mammuthus primigenius</i>	This study	L435
<i>Mammuthus primigenius</i>	This study	L436
<i>Mammuthus primigenius</i>	This study	L437
<i>Mammuthus primigenius</i>	This study	L450D
<i>Mammuthus primigenius</i>	This study	L451
<i>Mammuthus primigenius</i>	This study	L452
<i>Mammuthus primigenius</i>	This study	L453
<i>Mammuthus primigenius</i>	This study	L454
<i>Mammuthus primigenius</i>	This study	L457
<i>Mammuthus primigenius</i>	This study	L458
<i>Mammuthus primigenius</i>	This study	L459
<i>Mammuthus primigenius</i>	This study	L460
<i>Mammuthus primigenius</i>	This study	L461
<i>Mammuthus primigenius</i>	This study	L463
<i>Mammuthus primigenius</i>	This study	L464
<i>Mammuthus primigenius</i>	This study	L465D
<i>Mammuthus primigenius</i>	This study	L466
<i>Mammuthus primigenius</i>	This study	L468

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	L469
<i>Mammuthus primigenius</i>	This study	L471
<i>Mammuthus primigenius</i>	This study	L472
<i>Mammuthus primigenius</i>	This study	L473
<i>Mammuthus primigenius</i>	This study	L475
<i>Mammuthus primigenius</i>	This study	L476
<i>Mammuthus primigenius</i>	This study	L477D
<i>Mammuthus primigenius</i>	This study	L478D
<i>Mammuthus primigenius</i>	This study	L479D
<i>Mammuthus primigenius</i>	This study	L480
<i>Mammuthus primigenius</i>	This study	L481
<i>Mammuthus primigenius</i>	This study	L483
<i>Mammuthus primigenius</i>	This study	L484
<i>Mammuthus primigenius</i>	This study	L485
<i>Mammuthus primigenius</i>	This study	L486
<i>Mammuthus primigenius</i>	This study	L487
<i>Mammuthus primigenius</i>	This study	L488
<i>Mammuthus primigenius</i>	This study	L489
<i>Mammuthus primigenius</i>	This study	L491
<i>Mammuthus</i> sp.	This study	LD090
<i>Mammuthus</i> sp.	This study	LD091
<i>Mammuthus primigenius</i>	This study	M10
<i>Mammuthus primigenius</i>	This study	M11D
<i>Mammuthus primigenius</i>	This study	M12
<i>Mammuthus primigenius</i>	This study	M13
<i>Mammuthus primigenius</i>	This study	M14
<i>Mammuthus primigenius</i>	This study	M15
<i>Mammuthus primigenius</i>	This study	M16
<i>Mammuthus primigenius</i>	This study	M18
<i>Mammuthus primigenius</i>	This study	M19
<i>Mammuthus primigenius</i>	This study	M2
<i>Mammuthus primigenius</i>	This study	M20
<i>Mammuthus primigenius</i>	This study	M21
<i>Mammuthus primigenius</i>	This study	M22
<i>Mammuthus primigenius</i>	This study	M23
<i>Mammuthus primigenius</i>	This study	M25
<i>Mammuthus primigenius</i>	This study	M26
<i>Mammuthus primigenius</i>	This study	M27
<i>Mammuthus primigenius</i>	This study	M28
<i>Mammuthus primigenius</i>	This study	M29
<i>Mammuthus primigenius</i>	This study	M30
<i>Mammuthus primigenius</i>	This study	M31
<i>Mammuthus primigenius</i>	This study	M36
<i>Mammuthus primigenius</i>	This study	M40
<i>Mammuthus primigenius</i>	This study	M5
<i>Mammuthus primigenius</i>	This study	M7
<i>Mammuthus primigenius</i>	This study	M8
<i>Mammuthus primigenius</i>	This study	MD001

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	MD003
<i>Mammuthus primigenius</i>	This study	MD004
<i>Mammuthus primigenius</i>	This study	MD005
<i>Mammuthus primigenius</i>	This study	MD006
<i>Mammuthus primigenius</i>	This study	MD012
<i>Mammuthus primigenius</i>	This study	MD013
<i>Mammuthus primigenius</i>	This study	MD014
<i>Mammuthus primigenius</i>	This study	MD015
<i>Mammuthus primigenius</i>	This study	MD016
<i>Mammuthus primigenius</i>	This study	MD017
<i>Mammuthus primigenius</i>	This study	MD021
<i>Mammuthus primigenius</i>	This study	MD022
<i>Mammuthus primigenius</i>	This study	MD023
<i>Mammuthus primigenius</i>	This study	MD025
<i>Mammuthus primigenius</i>	This study	MD026
<i>Mammuthus primigenius</i>	This study	MD027
<i>Mammuthus primigenius</i>	This study	MD028
<i>Mammuthus primigenius</i>	This study	MD029
<i>Mammuthus primigenius</i>	This study	MD031
<i>Mammuthus primigenius</i>	This study	MD032
<i>Mammuthus primigenius</i>	This study	MD033
<i>Mammuthus primigenius</i>	This study	MD034
<i>Mammuthus primigenius</i>	This study	MD035
<i>Mammuthus primigenius</i>	This study	MD036
<i>Mammuthus primigenius</i>	This study	MD037
<i>Mammuthus primigenius</i>	This study	MD038
<i>Mammuthus primigenius</i>	This study	MD040
<i>Mammuthus primigenius</i>	This study	MD041
<i>Mammuthus primigenius</i>	This study	MD043
<i>Mammuthus primigenius</i>	This study	MD044
<i>Mammuthus primigenius</i>	This study	MD047
<i>Mammuthus primigenius</i>	This study	MD048
<i>Mammuthus primigenius</i>	This study	MD049
<i>Mammuthus primigenius</i>	This study	MD050
<i>Mammuthus primigenius</i>	This study	MD051
<i>Mammuthus primigenius</i>	This study	MD052
<i>Mammuthus primigenius</i>	This study	MD053
<i>Mammuthus primigenius</i>	This study	MD054
<i>Mammuthus primigenius</i>	This study	MD055
<i>Mammuthus primigenius</i>	This study	MD056
<i>Mammuthus primigenius</i>	This study	MD057
<i>Mammuthus primigenius</i>	This study	MD058
<i>Mammuthus primigenius</i>	This study	MD059
<i>Mammuthus primigenius</i>	This study	MD060
<i>Mammuthus primigenius</i>	This study	MD062
<i>Mammuthus primigenius</i>	This study	MD063
<i>Mammuthus primigenius</i>	This study	MD064
<i>Mammuthus primigenius</i>	This study	MD065

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	MD066
<i>Mammuthus primigenius</i>	This study	MD067
<i>Mammuthus primigenius</i>	This study	MD068
<i>Mammuthus primigenius</i>	This study	MD069
<i>Mammuthus primigenius</i>	This study	MD070
<i>Mammuthus primigenius</i>	This study	MD071
<i>Mammuthus primigenius</i>	This study	MD072
<i>Mammuthus primigenius</i>	This study	MD073
<i>Mammuthus primigenius</i>	This study	MD074
<i>Mammuthus primigenius</i>	This study	MD075
<i>Mammuthus primigenius</i>	This study	MD077
<i>Mammuthus primigenius</i>	This study	MD078
<i>Mammuthus primigenius</i>	This study	MD079
<i>Mammuthus primigenius</i>	This study	MD081
<i>Mammuthus primigenius</i>	This study	MD082
<i>Mammuthus primigenius</i>	This study	MD083
<i>Mammuthus primigenius</i>	This study	MD084
<i>Mammuthus primigenius</i>	This study	MD085
<i>Mammuthus primigenius</i>	This study	MD086
<i>Mammuthus primigenius</i>	This study	MD087
<i>Mammuthus primigenius</i>	This study	MD088
<i>Mammuthus primigenius</i>	This study	MD089
<i>Mammuthus primigenius</i>	This study	MD091
<i>Mammuthus primigenius</i>	This study	MD092
<i>Mammuthus primigenius</i>	This study	MD094
<i>Mammuthus primigenius</i>	This study	MD095
<i>Mammuthus primigenius</i>	This study	MD096
<i>Mammuthus primigenius</i>	This study	MD097
<i>Mammuthus primigenius</i>	This study	MD098
<i>Mammuthus primigenius</i>	This study	MD099
<i>Mammuthus primigenius</i>	This study	MD100
<i>Mammuthus primigenius</i>	This study	MD101
<i>Mammuthus primigenius</i>	This study	MD102
<i>Mammuthus primigenius</i>	This study	MD103
<i>Mammuthus primigenius</i>	This study	MD104
<i>Mammuthus primigenius</i>	This study	MD105
<i>Mammuthus primigenius</i>	This study	MD106
<i>Mammuthus primigenius</i>	This study	MD107
<i>Mammuthus primigenius</i>	This study	MD108
<i>Mammuthus primigenius</i>	This study	MD109
<i>Mammuthus primigenius</i>	This study	MD110
<i>Mammuthus primigenius</i>	This study	MD111
<i>Mammuthus primigenius</i>	This study	MD112
<i>Mammuthus primigenius</i>	This study	MD113
<i>Mammuthus primigenius</i>	This study	MD114
<i>Mammuthus primigenius</i>	This study	MD115
<i>Mammuthus primigenius</i>	This study	MD116
<i>Mammuthus primigenius</i>	This study	MD117

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	MD118
<i>Mammuthus primigenius</i>	This study	MD119
<i>Mammuthus primigenius</i>	This study	MD120
<i>Mammuthus primigenius</i>	This study	MD121
<i>Mammuthus primigenius</i>	This study	MD122
<i>Mammuthus primigenius</i>	This study	MD123
<i>Mammuthus primigenius</i>	This study	MD124
<i>Mammuthus primigenius</i>	This study	MD125
<i>Mammuthus primigenius</i>	This study	MD126
<i>Mammuthus primigenius</i>	This study	MD127
<i>Mammuthus primigenius</i>	This study	MD128
<i>Mammuthus primigenius</i>	This study	MD129
<i>Mammuthus primigenius</i>	This study	MD130
<i>Mammuthus primigenius</i>	This study	MD131
<i>Mammuthus primigenius</i>	This study	MD132
<i>Mammuthus primigenius</i>	This study	MD133
<i>Mammuthus primigenius</i>	This study	MD134
<i>Mammuthus primigenius</i>	This study	MD135
<i>Mammuthus primigenius</i>	This study	MD136
<i>Mammuthus primigenius</i>	This study	MD137
<i>Mammuthus primigenius</i>	This study	MD138
<i>Mammuthus primigenius</i>	This study	MD139
<i>Mammuthus primigenius</i>	This study	MD140
<i>Mammuthus primigenius</i>	This study	MD141
<i>Mammuthus primigenius</i>	This study	MD142
<i>Mammuthus primigenius</i>	This study	MD143
<i>Mammuthus primigenius</i>	This study	MD144
<i>Mammuthus primigenius</i>	This study	MD145
<i>Mammuthus primigenius</i>	This study	MD146
<i>Mammuthus primigenius</i>	This study	MD147
<i>Mammuthus primigenius</i>	This study	MD148
<i>Mammuthus primigenius</i>	This study	MD149
<i>Mammuthus primigenius</i>	This study	MD150
<i>Mammuthus primigenius</i>	This study	MD151
<i>Mammuthus primigenius</i>	This study	MD152
<i>Mammuthus primigenius</i>	This study	MD153
<i>Mammuthus primigenius</i>	This study	MD155
<i>Mammuthus primigenius</i>	This study	MD156
<i>Mammuthus primigenius</i>	This study	MD157
<i>Mammuthus primigenius</i>	This study	MD158
<i>Mammuthus primigenius</i>	This study	MD159
<i>Mammuthus primigenius</i>	This study	MD160
<i>Mammuthus primigenius</i>	This study	MD161
<i>Mammuthus primigenius</i>	This study	MD162
<i>Mammuthus primigenius</i>	This study	MD163
<i>Mammuthus primigenius</i>	This study	MD164
<i>Mammuthus primigenius</i>	This study	MD165
<i>Mammuthus primigenius</i>	This study	MD166

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	MD167
<i>Mammuthus primigenius</i>	This study	MD174
<i>Mammuthus primigenius</i>	This study	MD176
<i>Mammuthus primigenius</i>	This study	MD177
<i>Mammuthus primigenius</i>	This study	MD178
<i>Mammuthus primigenius</i>	This study	MD179
<i>Mammuthus primigenius</i>	This study	MD180
<i>Mammuthus primigenius</i>	This study	MD181
<i>Mammuthus primigenius</i>	This study	MD182
<i>Mammuthus primigenius</i>	This study	MD183
<i>Mammuthus primigenius</i>	This study	MD184
<i>Mammuthus primigenius</i>	This study	MD185
<i>Mammuthus primigenius</i>	This study	MD189
<i>Mammuthus primigenius</i>	This study	MD191
<i>Mammuthus primigenius</i>	This study	MD192
<i>Mammuthus primigenius</i>	This study	MD194
<i>Mammuthus primigenius</i>	This study	MD197
<i>Mammuthus primigenius</i>	This study	MD199
<i>Mammuthus primigenius</i>	This study	MD200
<i>Mammuthus primigenius</i>	This study	MD201
<i>Mammuthus primigenius</i>	This study	MD202
<i>Mammuthus primigenius</i>	This study	MD203
<i>Mammuthus primigenius</i>	This study	MD204
<i>Mammuthus primigenius</i>	This study	MD206
<i>Mammuthus primigenius</i>	This study	MD207
<i>Mammuthus primigenius</i>	This study	MD208
<i>Mammuthus primigenius</i>	This study	MD209
<i>Mammuthus primigenius</i>	This study	MD210
<i>Mammuthus primigenius</i>	This study	MD211
<i>Mammuthus primigenius</i>	This study	MD212
<i>Mammuthus primigenius</i>	This study	MD213
<i>Mammuthus primigenius</i>	This study	MD214
<i>Mammuthus primigenius</i>	This study	MD215
<i>Mammuthus primigenius</i>	This study	MD216
<i>Mammuthus primigenius</i>	This study	MD217
<i>Mammuthus trogontherii</i>	This study	MD218
<i>Mammuthus trogontherii</i>	This study	MD219
<i>Mammuthus trogontherii</i>	This study	MD220
<i>Mammuthus trogontherii</i>	This study	MD221
<i>Mammuthus trogontherii</i>	This study	MD222
<i>Mammuthus trogontherii</i>	This study	MD223
<i>Mammuthus trogontherii</i>	This study	MD224
<i>Mammuthus trogontherii</i>	This study	MD225
<i>Mammuthus trogontherii</i>	This study	MD226
<i>Mammuthus primigenius</i>	This study	MD227
<i>Mammuthus primigenius</i>	This study	MD228
<i>Mammuthus primigenius</i>	This study	MD229
<i>Mammuthus trogontherii</i>	This study	MD230

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus trogontherii</i>	This study	MD231
<i>Mammuthus trogontherii</i>	This study	MD232
<i>Mammuthus trogontherii</i>	This study	MD233
<i>Mammuthus trogontherii</i>	This study	MD234
<i>Mammuthus trogontherii</i>	This study	MD235
<i>Mammuthus trogontherii</i>	This study	MD236
<i>Mammuthus trogontherii</i>	This study	MD237
<i>Mammuthus trogontherii</i>	This study	MD238
<i>Mammuthus trogontherii</i>	This study	MD239
<i>Mammuthus trogontherii</i>	This study	MD240
<i>Mammuthus trogontherii</i>	This study	MD241
<i>Mammuthus trogontherii</i>	This study	MD242
<i>Mammuthus trogontherii</i>	This study	MD243
<i>Mammuthus trogontherii</i>	This study	MD244
<i>Mammuthus trogontherii</i>	This study	MD245
<i>Mammuthus trogontherii</i>	This study	MD246
<i>Mammuthus</i> sp.	This study	MD247
<i>Mammuthus primigenius</i>	This study	MH322
<i>Mammuthus primigenius</i>	This study	P001
<i>Mammuthus primigenius</i>	This study	P002
<i>Mammuthus primigenius</i>	This study	P003
<i>Mammuthus primigenius</i>	This study	P008
<i>Mammuthus primigenius</i>	This study	P009
<i>Mammuthus primigenius</i>	This study	P011
<i>Mammuthus primigenius</i>	This study	P012
<i>Mammuthus primigenius</i>	This study	P013
<i>Mammuthus primigenius</i>	This study	P016
<i>Mammuthus primigenius</i>	This study	P017
<i>Mammuthus primigenius</i>	This study	P020
<i>Mammuthus primigenius</i>	This study	P021
<i>Mammuthus primigenius</i>	This study	P022
<i>Mammuthus primigenius</i>	This study	P023
<i>Mammuthus primigenius</i>	This study	P024
<i>Mammuthus primigenius</i>	This study	P026
<i>Mammuthus primigenius</i>	This study	P027
<i>Mammuthus primigenius</i>	This study	P028
<i>Mammuthus primigenius</i>	This study	P029
<i>Mammuthus primigenius</i>	This study	P030
<i>Mammuthus</i> sp.	This study	P031
<i>Mammuthus</i> sp.	This study	P032
<i>Mammuthus</i> sp.	This study	P034
<i>Mammuthus trogontherii</i>	This study	P035
<i>Mammuthus</i> sp.	This study	P036
<i>Mammuthus</i> sp.	This study	P038
<i>Mammuthus</i> sp.	This study	P039
<i>Mammuthus</i> sp.	This study	P040
<i>Mammuthus</i> sp.	This study	P041
<i>Mammuthus</i> sp.	This study	P042

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<i>Mammuthus primigenius</i>	This study	P045
<i>Mammuthus</i> sp.	This study	P046
<i>Mammuthus</i> sp.	This study	P048
<i>Mammuthus</i> sp.	This study	P049
<i>Mammuthus primigenius</i>	This study	P052
<i>Loxodonta africana</i>	This study	P073
<i>Loxodonta africana</i>	This study	P074
<i>Loxodonta africana</i>	This study	P075
<i>Mammuthus</i> sp.	This study	Mammuthus-Yuka
<i>Mammuthus</i> sp.	Van der Valk et al. ⁶⁷	L082
<i>Mammuthus</i> sp.	Van der Valk et al. ⁶⁷	L286
<i>Mammuthus trogontherii</i>	Van der Valk et al. ⁶⁷	P033
<i>Mammuthus trogontherii</i>	Van der Valk et al. ⁶⁷	P037
<i>Mammuthus primigenius</i>	Van der Valk et al. ⁶⁷	P043
<i>Mammuthus primigenius</i>	Van der Valk et al. ⁶⁸	L163
<i>Mammuthus primigenius</i>	Van der Valk et al. ⁶⁸	M6

Chemicals, peptides, and recombinant proteins

USER enzyme	New England Biolabs	NEB #M5508
AccuPrime reaction mix	Life Technologies	Cat #12344040
AccuPrime Pfx DNA polymerase	Life Technologies	Cat #12344024
EDTA	ThermoFisher Scientific	Cat #15575020
UREA	VWR	Cat #443874G
Proteinase K	VWR	Cat #1.24568.0100
Tango Buffer (10X)	ThermoFisher Scientific	Cat #BY5
ATP (100mM)	ThermoFisher Scientific	Cat #R0441
T4 Polynucleotide Kinase (10U/ul)	ThermoFisher Scientific	Cat #EK0032
T4 DNA Polymerase (5U/ul)	ThermoFisher Scientific	Cat #EP0062
T4 DNA Ligase (5U/ul)	ThermoFisher Scientific	Cat #EL0011
Bst Polymerase, LF (8U/ul)	New England Biolabs	Cat #M0275S
Maxima SYBR Green/ROX qPCR Master Mix (2X)	ThermoFisher Scientific	Cat #11873913
Agencourt AMPure XP beads	Beckman Coulter	Cat #10136224

Critical commercial assays

MinElute PCR purification kit	QIAGEN	Cat #28115
QiaQuick PCR purification kit	QIAGEN	Cat #28106
Agilent High Sensitivity kit	Agilent	Cat #5067-4626

Deposited data

Raw sequencing data	This study	ENA Project Number PRJEB78615
Bioinformatic codes	This study	https://github.com/BenjaminGuinet/Mammuth_Metagenomics

Software and algorithms

Amber v1	Dolenz et al. ²⁵	https://github.com/tvandervalk/AMBER
ANGSD v.0.940	Durvasula et al. ²⁶	https://github.com/ANGSD/angsd
BamUtil trimBam v1.0.15	Jun et al. ⁴⁰	https://github.com/statgen/bamUtil
Bowtie v.2.5.1	Langmead and Salzberg ⁴⁴	http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
BWA aln v.0.7.8	Li and Durbin ⁴⁶	http://bio-bwa.sourceforge.net
Fastp v.0.20.0	Chen et al. ¹³	https://github.com/OpenGene/fastp
GTDB Database v.17/08/2023	Parks et al. ⁶¹	https://gtdb.ecogenomic.org/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Iqtree v.2.3.5	Nguyen et al. ⁵⁵	http://www.iqtree.org/
Kraken2 v.2.1.2	Wood et al. ⁶⁹	https://github.com/DerrickWood/kraken2
Mmseqs2 v15-6f452	Steinegger et al. ⁷⁰	https://github.com/soedinglab/MMseqs2
ModelFinder v1	Kalyaanamoorthy et al. ⁴¹	http://www.iqtree.org/ModelFinder/
NCBI datasets v17.1.0	O'Leary et al. ⁵⁶	https://github.com/ncbi/datasets
PanACoTA v. 1.4.2-dev1	Perrin et al. ⁶²	https://github.com/gem-pasteur/PanACoTA
PathPhynder v1	Maertiniario et al. ⁵⁰	https://github.com/SayakaMiura/PathFinder
PMDTools v0.6	Skoglund et al. ⁷¹	https://github.com/pontusssk/PMDtools
PySam	PySam Developers	https://github.com/pysam-developers/pysam
R v4.4.1-cpeGNU-23.12	R Core Team	https://www.R-project.org
RAxML EPA-ng v0.3.8	Barbera et al. ²	https://github.com/pierrebarbera/epa-ng
VFDB v.08/02/2024	Chen et al. ¹²	https://www.mgc.ac.cn/VFs/
VirulentPred v.2.0	Sharma et al. ⁷²	https://bioinfo.icgeb.res.in/virulent/index.html

METHOD DETAILS

Sample collection and processing

A total of 483 mammoth samples, of which 43 were previously published^{2–5,70,73,74} and 440 newly sequenced and unpublished samples from six Proboscidea species (*M. trogontherii*, *M. primigenius*, *M. imperator*, *M. columbi*, *P. antiquus*, *M. americanum*) were screened for the presence of ancient microbial DNA (see Figure S2 for full pipeline abstract details). For the unpublished woolly mammoth samples DNA extractions and library preparations were performed according to standard ancient DNA practices at the dedicated ancient DNA lab facilities of either the Swedish Museum of Natural History or the Centre for Palaeogenetics, both located in Stockholm, Sweden. Briefly, 50–200 mg of bone or tooth powder was collected using a Dremel drill. DNA extractions were carried out using the silica column protocol as described in Dehasque et al.⁷⁵ After overnight digestion, the extraction protocol continued from day two as specified in Dehasque et al. Double-stranded sequencing libraries (except for one of the *Adyca* (sample P033) extractions made from single-stranded libraries) were then prepared following the protocol by Meyer and Kircher⁷⁶, including treatment with either 3 or 6 μ L of USER (New England Biolabs), as described in Dehasque et al. The USER enzyme, a mixture of uracil-DNA-glycosylase (UDG) and endonuclease VIII (endoVIII), exercises uracil bases incorporated due to *post mortem* damage, except at CpG sites. The majority of the species analyzed were *M. primigenius* (n=414) and *M. trogontherii* (n=34). Additionally, we included one sample from a *M. imperator* (n=1), as well as published Proboscidea sequence data from *M. columbi* (n=1), *Mammuth americanum* (n=2), *Palaeoloxodon antiquus* (n=3) and historical African elephant samples (*Loxodonta africana*) (n=3), but for none of these species we identified authenticated ancient microbes and were thus not included further in this study. The remaining samples (n=36) are not assigned to a specific species taxonomy but all of them represented *Mammuthus* species (see Table S1 for all metadata). The samples were collected from various locations ranging from North America and Britain to Siberia (Figure S1). The ages of the samples varied, with recent (4000–10,000 years ago) *M. primigenius* specimens from Wrangel Island, and *M. trogontherii* samples dating back to >1 million years ago from Eastern Siberia (Figure S1). The majority of samples came from mammoth tusks and molars, with some samples obtained from various bones (Table S1) and one skin sample.

Metagenomic screening

In this study we aimed to describe ancient microbiota present in mammoth samples. We specifically categorized microbes as organisms belonging to the prokaryotic domains of Bacteria and Archaea. As a first step we aimed to remove all sequence reads that are from either the mammoth genome or human contamination. To do so, sequence adapters were removed, and paired-end reads were merged and deduplicated using *fastp* v.0.20.0⁶⁷ so that long reads that are more likely to correspond to modern contaminant DNA are excluded. Additionally, reads below 30bp were removed from the dataset with *fastp* as these often lead to spurious mappings. The merged reads with a minimum length of 30bp were then mapped to a concatenated reference of the human (hg19) and Asian elephant reference (RefSeq: GCF_024166365.1) and mammoth mitogenome (GenBank: NC_007596.2), using *BWA-aln* v.0.7.8 with deactivated seeding (*-l* 16,500), allowing for more substitutions (*-n* 0.01) and up to two gaps (*-o* 2).⁷² Then, only unmapped reads were further analyzed in the next steps. As a first classification step, we classified the unmapped reads using *Kraken2* (v2.1.2) (default parameters)⁷⁷ against an inhouse build GTDB database (built 17/08/2023),⁶⁸ consisting of contigs of isolate genomes, metagenome-assembled genomes (MAGs), and single-amplified genomes, comprising 394,932 bacterial and 7,777 archaeal genomes organized into 80,789 bacterial and 4,416 archaeal species clusters.

We chose the GTDB database (created on 17/08/2023) because we found it to be more comprehensive than the traditional NCBI RefSeq database (created on 30/05/2024) (see [Data S1](#) – Section A for more details).

Microbial candidate species, containing over 1,000 distinct k-mers (specific to that particular species) and having over 200 classified reads, were selected for further analysis. *Kraken2* read classification is prone to a high false positive error rate due to misaligned reads, where reads from a species not present in the database are incorrectly assigned to the conserved sequence regions of another related species in the database. In such cases it is common that only a fraction of the database reference genome has assigned reads. We implemented quality-control filters to minimize false-positive species identifications by competitively mapping the unclassified reads against all the genomes of the *Kraken2* detected microbial species using *bowtie2* v2.5.1 with the –very-sensitive setting.

To evaluate whether this method effectively maps reads to distantly related reference genomes, we tested the *bowtie2* parameters used in this study for their ability to align such reads. This is particularly relevant because most candidate microbes in our dataset are expected to be distantly related bacteria, due to both the incompleteness of current bacterial databases and the absence of any described bacterial species from mammoths. We therefore downloaded various *Yersinia* genomes, split them into 50bp reads, and mapped these reads against the *Yersinia pestis* genome. The results indicated that when the reference genome is closely related (>98% similarity), we could still map over 90% of the reads with our parameters (see [Data S1](#) – Section B for details). However, when the reference genome is more distant (<96% similarity), only a small proportion of the reads could be mapped, likely corresponding to conserved regions between the two species. These results suggest that our approach can recover some of the reads but may not capture all of them.

To keep the *bowtie2* indexing within computational limits one reference microbial genome per genus was used, which was selected based on the most abundant *Kraken2* classified species within each genus. Although this method can alleviate computational requirements it may introduce biases where multiple species from the same genus are present in a sample. In such cases, a multimodal distribution of the percentage identity of the reads mapped against the given reference genome is observed. Therefore, in subsequent steps, particularly in phylogenetic inferences for the animal-associated bacteria, we ensured to include only species where the reads showed a unimodal read percentage identity distribution ([Data S4](#)).

Next, we used the Python package Pysam (<https://github.com/pysam-developers/pysam>) to assess whether the coverage of the reads mapped to each reference genome was uniformly distributed (evenness of coverage). We generated 100 windows per reference microbial genome and computed the proportion of windows with a mean depth of at least 0.01X the same threshold, as previously used in other ancient metagenomic pipeline.⁷⁸ Only reference genomes with at least 90% of windows above this threshold were considered to be evenly covered. We also calculated evenness of coverage with more stringent values using thresholds of 0.05 and 0.1 ([Table S12](#)) however, this approach was only suitable for high-coverage bacterial genomes.

Contaminating assessment

Ancient microbial samples typically contain a mixture of DNA derived from microbes present both before and after the host organism's death. This DNA may originate from several sources, including the host's endogenous microbiota, pathogens potentially responsible for disease (e.g., *Yersinia pestis* in tooth pulp), and environmental microorganisms such as soil bacteria involved in decomposition. Additionally, modern contamination can be introduced through sample handling (e.g., human skin-associated microbes), storage conditions, and laboratory procedures (e.g., reagents contaminated with vector-derived DNA).⁷⁹ To identify and remove putative modern contaminants, we applied a two-step read classification approach to 43 independent laboratory blanks processed in the same facility where most of our samples were handled. To account for potential sediment-derived contamination, we also analyzed a large metagenomic dataset (771 million reads) from Middle and Late Pleistocene sediments collected from the Batagay megaslump in East Siberia (SRA project PRJEB43506, downloaded on 2023-12-09).¹⁶ This region and time period are relevant, as several of our mammoth samples originate from the same context. To better understand the differences between the sediment samples and the mammoth samples, we computed a distance dendrogram using *MASH* v2.3.¹⁵ This dendrogram provides insights into the compositional relationships between the various read sources. Overall, most mammoth samples exhibited similar read compositions among themselves, and a subset of these samples also shared similarities with the sediment samples. However, the laboratory blanks displayed a distinctly different composition ([Data S1](#) – Section C). This sediment dataset was used to identify environmental microbes likely to have colonized mammoth remains *post mortem* and therefore not directly associated with the host. Candidate taxa were flagged as potential contaminants if reads assigned to them were evenly distributed across reference genomes defined as coverage over 90% across 100 bp windows. Such candidates were subjected to further phylogenetic analysis to confirm their origin. In the laboratory blanks, we consistently detected high abundances of bacterial genera such as *Burkholderia*, *Novosphingobium*, *Bradyrhizobium*, *Flavobacterium*, and *Curvibacter* ([Figure 1A](#)). Given their frequent occurrence, targeted depletion of these taxa prior to sequencing may be a cost-effective strategy to reduce the number of non-informative reads.

Authentication of ancient DNA

The authenticity of candidate ancient microbial reads was assessed using a custom scoring system designed to evaluate the genomic distribution of reads and post mortem DNA damage. The scoring system was based on five criteria, grouped into two categories:

Presence/Absence Evaluation:

- **(+1)** Evenness of coverage: >90% of 100 bp windows are covered by mapped reads, indicating a broad distribution rather than mapping to conserved or repetitive regions only.

- **(+1)** Short read lengths: Mean read length <70 bp, consistent with high DNA fragmentation.
- **(+1)** Damage signal: $\geq 8\%$ of reads with a *post mortem* damage (PMD) score >1.5 at CpG sites.

DNA Damage Evaluation:

- **(+1)** Terminal damage: $\geq 5\%$ C-to-T substitutions at the first position of reads (CpG or non-CpG sites).
- **(+1)** Damage gradient: Higher average damage in the first five bases than in the next 25 bases, with fewer than 10 positions where the DNA damage is >200% the read average (due to stochasticity).

We deliberately excluded sequence genetic similarity to known genomes from the scoring criteria, as many of the microbes recovered from our samples are expected to only be distantly related to the reference sequences in GTDB (see Figure 1C). After scoring, we retained only those taxa with a minimum score of 4, an evenness of coverage >50%, and those showing evidence of DNA damage based on at least one of the two damage criteria. To validate this scoring system, we used *Yersinia pestis*, the causative agent of plague and previously identified in an ancient human sample (Gökhem2 (EMBL-EBI: SAMEA2482233) from Northern Europe⁸⁰ as a positive control. Our pipeline resulted in an authenticity score of 5/5 for this sample.

Considerations of UDG treatment and residual damage

Most mammoth samples in this study underwent partial or full uracil-DNA glycosylase (UDG) treatment, which reduces typical aDNA damage signatures (see Figure 1E). However, in 53 samples, UDG treatment was partial or incomplete, allowing residual C-to-T substitutions to persist at the terminal ends of reads (Table S3). For example, in a 1.1-million-year-old steppe mammoth sample (Figure 2), both UDG-treated and untreated libraries showed differential damage profiles, with UDG-treated reads exhibiting damage restricted to only the first few bases. Similarly, in *Y. pestis* control experiments (Figure S5), this pattern was reproduced, confirming the scoring system's sensitivity to authentic damage signals.

Final selection of ancient microbial candidates

From all taxonomically assigned microbial candidates in mammoth samples, 2,740 met the initial authenticity thresholds. We then visually inspected damage profiles for each, categorizing them into three groups based on the strength and clarity of their damage signals (see Figure S6):

- “Signal”: Clear C-to-T or CpG-to-TpG damage at read termini, minimal mid-read damage, and low stochasticity (Figure S6A).
- “Intermediate”: Damage profiles consistent with aDNA, but with some variability in damage gradient or stochasticity (Figure S6B).
- “No Signal”: Lack of consistent terminal damage patterns, with high stochasticity across read length (Figure S6C).

Ultimately, 310 microbial species exhibited either intermediate or clear damage signatures. These were detected across 105 mammoth samples. The majority of candidates ($n=2,430$) were classified as “no signal,” while 221 were “intermediate” and 89 showed clear aDNA damage. The relatively small proportion of taxa with authentic aDNA signals is likely due to the evolutionary divergence between the recovered reads and available reference genomes (Figure S5). We observed that as the genetic distance to the reference increased, characteristic damage patterns became increasingly stochastic and less pronounced. For instance, mapping ancient *Y. pestis* reads to more distantly related references resulted in reduced and more erratic damage profiles (Figure S5). We also found that sequencing depth positively correlated with the clarity of damage profiles, with deeper sequencing yielding smoother and more reliable damage signals (Pearson's $r = 0.66$, $p < 0.0001$, Figure S5). All mapping statistics and damage metrics are provided in Table S2.

Ecological niche analysis of microbes

To identify the putative ecological niche of the organism (e.g., environmental or animal), we relied on the Microbe Atlas project (MAP).²⁷ Using this database for each reference microbial species, we searched at the lowest level of taxonomy available (species, genus or family) and examined the fraction of specimens from the same taxonomic rank in MAP associated with ‘animal’, ‘aquatic’, ‘soil’, or ‘plant’. If the proportion of related specimens associated with animals exceeded 90%, we classified that microbial genus and the detected ancient microbes as being of likely animal associated. If more than 50% but less than 90% of the samples were assigned to “animal”, we classified the microbe as “animal/environmental”. Otherwise, we classified it as “environmental” (see Table S4). If the reference microbial species or genus was not present in the MAP database, we relied on the described isolation sources in the literature.

Visualization and plot analysis

For all of the microbial candidates remaining after the above filtering, we build a PDF plot consisting of six subgraphs, illustrating DNA damage patterns (mismatch plot, PMDscore distribution), read characteristics (read-length distribution, identity score), and evenness coverage distribution over the reference genome. Furthermore, a table was generated containing various information, such as the tissue type, sequencing location, age of the sample, number of other mammoth samples containing the same microbial species, and additional relevant data. The script was made using python along with multiple python libraries and is available on GitHub (https://github.com/BenjaminGuinet/DamagePlots_from_BAM). Each microbe plot was then visually inspected to ensure it was a reliable ancient candidate.

Phylogenetic inference

To investigate the phylogenetic relationships between ancient and modern microbial strains, we inferred phylogenetic trees using multiple independent methods (see [Figure S2](#) for full abstract details). The process started by aligning reference genomes from modern bacteria in the same genus as the detected ancient microbes in order to capture the maximum species diversity within the clade (assemblies were downloaded along with their genome metadata using NCBI datasets v17.1.0⁸¹ and a custom script (https://github.com/BenjaminGuinet/Extract_NCBI_assembly_metadata)). Only reads with a mapping quality of 30 or higher were used to minimize the inclusion of spuriously mapped reads (see the nucleotide identity distribution and edit distance distribution of these reads in [Data S4](#)). We then used *PanACoTA* v1.4.1²⁹ to identify single-copy orthologous genes across the different species and strains (annotate: default; pangenome: -i 0.7, cluster mode 1; corepers: -t 0.3). Next, using *PanACoTA*, we performed *de novo* alignment for each gene family/orthologous gene separately (align: default) and then concatenated these alignments to create a final alignment consisting solely of core/persistent microbial genes within the genus. To incorporate the ancient microbial genomes, we called consensus sequences using *ANGSD* v0.940⁷¹ (-doFasta 2 -doCounts 1 -minMapQ 30 -setMinDepth 2) after mapping the reads with *bowtie2* v2.5.2⁸² (-very-sensitive, retaining only reads with an edit distance < 3). Mapping was systematically done using as the reference the aligned parts of the modern bacteria genome that had the most classified unique k-mers in the *Kraken2* analysis. We included all mammoth ancient candidates where the mapped reads along the reference microbe genome showed at least 50% evenness of coverage (uniformness with which sequenced reads are distributed across the reference genome as described in⁷⁸). All gene alignments were then merged using a custom script (https://github.com/BenjaminGuinet/Mammoth_Metagenomics/blob/main/4-Phylogenetic_analysis/Merge_genes_to_MSA.py) to keep partition structure information. Phylogenetic trees on this final alignment were then inferred using a Maximum-likelihood framework (ML) and *ModelFinder*⁶⁹ was used to identify models for each ORF partition (-MFP option), both implemented in *IQ-TREE* v2.2.2.6.⁸³ The edge-linked partitioned model (-spp option), which allows each gene to have its own evolutionary rate, was chosen for tree reconstructions. Ultra-fast bootstrap⁸⁴ and SH-aLRT (options -bb 1000 and -alrt 1000) were computed to examine node supports for focal relationships using the ML method. Additionally, to reduce the risk of overestimating branch supports due to model violations, we used the command "-bnni".

Phylogenetic robustness analysis

The six newly identified mammoth associated clades in this study have an average genome coverage of 0.78X, with many of the samples having less than 20% genomic breadth coverage, making accurate phylogenetic placements challenging. Therefore, the monophyletic clustering of mammoth bacteria could be an artifact of shared missing data rather than a robust signal calculated on the aligned data. To test the robustness of the phylogenetic placement with low coverage data, we performed multiple phylogenetic evaluations.

Testing with phylogenetic placement algorithms

We evaluated the phylogenetic placement of mammoth bacterial sequences using the Evolutionary Placement Algorithm *EPA-ng*, which is integrated into *RAXML EPA-ng* v0.3.8.⁸⁵ This algorithm determines the optimal insertion position (termed placement) for each mammoth bacteria query sequence individually and independently on a fixed reference tree which is composed of the full modern genomes previously aligned with *PanACoTA*. The "optimal" position is identified by the likelihood score of the reference tree after adding one query sequence. The reference tree remains static, meaning each query is inserted into the same reference tree. Thus, rather than iteratively expanding the tree with each query sequence, the queries are mapped one-by-one to the best-scoring branches (insertion positions) in the reference tree. The same alignments as those in the main [Figure 4](#) were used, but removing the mammoth sequences from them prior to building the reference trees. Only reads with mapping quality ≥ 30 and min depth 2 were kept. We used *Gappa* v0.8.5⁸⁶ (specifically the *gappa* examine assign function) to summarize placement results and calculate the likelihood weight ratio (LWR) for each query's placements. The LWR quantifies the relative likelihood of a given placement compared to alternatives, with values close to 1 indicating high confidence at that node. Summing LWRs at various taxonomic levels (e.g., species, clade, family), allowed us to assess the most probable phylogenetic affiliation of each query sequence. For example, if placements were consistent within a clade but ambiguous at the species level, the LWR reflected this uncertainty. [Table S6](#) summarizes these LWRs across taxonomic levels.

For the genus *Erysipelothrix*, taxonomic placement was consistent across all mammoth samples, with LWR supporting its assignment within the Erysipelotrichaceae family confidently placed close to *Erysipelothrix tonsillarum*. The lowest support for *E. tonsillarum* assignment was observed in *Mammuthus*-MD024 (LWR = 0.9783), while the highest was recorded in *Mammuthus*-FK001 and other specimens (LWR = 0.9998). The placements were in agreement with the main phylogeny ([Figure 2](#)).

In the case of *Streptococcus*, the results were also concordant with the main phylogenetic result in [Figure 2](#). *Mammuthus*-MD228 showed strong assignment to *Streptococcus mutans* (LWR = 0.8214–0.8253), while *Mammuthus*-L426, *Mammuthus*-L423, *Mammuthus*-L422, and *Mammuthus*-L414 were confidently placed within *Streptococcus devriesei* (LWR = 0.8694). *Mammuthus*-M25 presented a low confidence toward a placement close to *Streptococcus mutans*.

For the *Actinobacillus* genus, all analyzed mammoth samples exhibited strong support for taxonomic placement close to *Actinobacillus* sp. GY-402. The lowest support was observed in *Mammuthus*-MD228 (LWR = 0.9028), whereas other mammoth such as *Mammuthus*-M40, *Mammuthus*-L389, and *Mammuthus*-FK012 had maximal assignment scores (LWR = 0.9998). These placements were fully consistent with the main phylogenetic results ([Figure 2](#)).

For the *Pasteurella* genus, all analyzed mammoth sample reads exhibited strong support for taxonomic placement close to Bisgaard Taxon 45 that was isolated from an African elephant. The lowest support was observed in *Mammuthus*-MD204 (LWR = 0.9702), whereas *Mammuthus*-FK003, had the highest maximal assignment score (LWR = 0.9795). These placements were fully consistent with the main phylogenetic results (Figure 2).

Overall, the taxonomic assignments obtained through *EPA-ng* and *Gappa* v0.9.0² demonstrated robust placement of all microbial sequences (except for the Mammoth M25) within established bacterial lineages, supporting their phylogenetic validity in the context of ancient mammoth microbiomes.

To further examine the SNPs differences between modern and ancient mammoth bacteria, we also employed an additional placement algorithm, *pathPhynder* V1⁸⁷ which assigns SNPs to tree branches and determines the optimal path for each sample within the tree using the same reads as those that were used for the *EPA-ng* analysis. Overall, the best placements identified by both *pathPhynder* and *EPA-ng* gave similar results, aligning with the main phylogeny shown in Figure 4 (Table S6).

One mammoth bacterium at a time with stringent filters

After confirming the placement of the mammoth sequences within the phylogenies including all modern species available for a given genus using the three independent methods (*EPA-ng*, *pathPhynder* and the maximum likelihoods in *IQ-TREE*), we aimed to build phylogenies using the maximize the number of sites covered by mammoth reads. To achieve this, we refined our phylogenetic analysis by selecting a subset of each phylogeny. This subset included only the closest assigned modern bacterial species for each genus, along with a few outgroup species within the same genus. This approach was designed to assign more gene families in the final sequence alignment constructed with *PanACoTA*. By doing so, we increased our chances of including additional mammoth-associated bacterial sites in the analysis. This allowed us to test more stringent parameters, such as site depth coverage or the number of sites shared between mammoth sequences.

To build the following phylogenies, the same methods using *PanACoTA* were used to align the sequences and *IQ-TREE* to infer the phylogenies. We then tested for the placement of each mammoth bacteria independently from each other with more stringent parameters. Indeed, since most microbial candidates exhibited very low breadth of coverage, there was a concern that ancient microbial lineages might cluster together within the phylogenetic tree due to artificial attraction. To address this, we systematically re-ran the phylogenetic analysis by introducing one mammoth-derived bacterial sequence at a time. Each alignment consisted of modern bacteria along with one mammoth sequence. We retained only the positions covered by the mammoth sequence, discarding all mammoth N positions. The same phylogenies were built across two different conditions:

1. Mapping quality (MAPQ) ≥ 30 and edit distance < 3 (Table S5C)
2. Mapping quality (MAPQ) ≥ 30 , mean depth coverage > 1 and edit distance < 3 (Table S5D)

Overall, one exception was observed:

- The *S. mutans* MD228 sequence shifted to an outgroup position between the *S. mutans* and *S. troglodytae* clades, whereas it was previously nested closer to the *S. mutans* clade.

Although these exceptions altered the closest microbial species identified, the changes were minor, with the new placements still being phylogenetically proximal to the original species. These shifts suggest that additional mammoth sequences might either help refine the phylogenetic resolution or introduce biases. Addressing this uncertainty will require deeper sequencing of these three specific samples. Until then, the placement of both M25 and MD228 in the primary phylogeny (Figure 2) should be interpreted cautiously, particularly concerning their positions relative to other mammoth sequences.

Only selecting shared sites between mammoths

To reliably assign species from ancient genomic datasets, especially for newly identified species linked to mammoth remains, high-coverage datasets are crucial. However, the six newly identified clades in this study are based on low-quality genomic data, with an average genome coverage of 0.78X and most samples having less than 20% genomic breadth coverage, making accurate species assignment difficult. To ensure that the phylogenetic placement of candidate ancient mammoth bacterial sequences relative to modern bacteria is not biased by incomplete genome coverage or other factors, it is important to consider that the clustering of mammoth bacteria could be an artifact due to shared missing data or very short, shared regions rather than a robust signal from the aligned data. Only in the case of ancient strains related to Bisgaard Taxon45 and *Streptococcus devriesei* do the mammoth strains have sufficiently covered sites to exclude clustering artifacts due to shared missing data (Table S12). When running the phylogeny with the same settings using only shared sites between mammoths in these two cases, a monophyletic clade among mammoth bacteria is maintained, although the bootstrap supports for *Pasteurella*-like mammoth bacteria are not statistically robust (Table S5B).

Divergence analysis of ancient microbial sequences

In several cases, we did not observe the expected branch shortening for the ancient microbial sequences recovered from the mammoth samples (Figure 4). Such shortening is typically anticipated in ancient DNA due to fewer mutations having accumulated over time relative to modern microbial lineages. However, accurately estimating branch length in our dataset is challenging for several

reasons. First, the low coverage of the ancient sequences, despite applying a depth threshold of $DP \geq 2$, means that sequencing or PCR errors introduced during library preparation may be misinterpreted as genuine variants. Additionally, the lack of observable shortening could also stem from comparing evolutionarily distant species, where deep divergence times may mask the relative difference in mutation accumulation between ancient and modern samples. To find further support for the identification of the ancient microbial sequences, we hypothesized that sequences from putative ancient microbes within a clade exhibit lower dissimilarity to the closest outgroup of the clade compared to modern bacteria (Figure 6A). This expectation is based on the assumption that less evolutionary time (compared to an outgroup) has elapsed for ancient microbial candidates, resulting in fewer derived mutations than modern representatives of the same clade. To compute dissimilarities between sequences, pairwise sequence distances were calculated from the gene alignments that were made using *PanACoTA* with a custom Python script (https://github.com/BenjaminGuinet/Mammuth_Metagenomics/blob/main/4-Phylogenetic_analysis/Calculate_MSA_percID.py). The method computes pairwise identity for each pair of sequences by comparing aligned positions, excluding gaps (-), ambiguous characters (N, X) and sites not covered by mammoth bacterial sequences.

Virulent gene identification

To investigate the pathogenic potential of the detected ancient bacteria we analyzed all the reads that mapped across the mammoth samples corresponding to the same clade in the phylogenetic tree together. We first mapped the reads using *bowtie2* (v2.5.2 –very-sensitive)⁸² and a consensus fasta sequence was then made using *ANGSD*, selecting the majority base at each site with aligned reads (v0.940 -doFasta 2 -doCounts 2).⁷¹ For each of these consensus assemblies we then screened for the presence of putative virulence genes by blasting the assemblies against the bacteria Virulence Factor DataBase (VFDB)⁸⁸ using *Mmseqs2 search* (v15-6f452) with min bitscore=50.⁸⁹ In total, we obtained 136 virulent gene hits. We then translated all the ancient putative virulent gene sequences into amino acids and ran a machine learning program (VirulentPred v2.0³⁰). VirulentPred v2.0 was trained on data from 32 different genera of pathogenic bacteria. It uses this training to identify patterns and features in protein sequences that are associated with virulence, which is the ability of bacteria to cause disease. By applying VirulentPred v2.0 to our translated protein sequences, we aimed to predict whether these ancient genes are likely to encode virulence factors. Conservatively, we then selected only the ancient microbial virulence proteins that covered at least 90% of the corresponding protein in the database. In total, 15 of these proteins were considered as reliable hits (see Table S9).

Microbial species divergence analysis of *Erysipelothrix*

An ancient microbe closely related to *Erysipelothrix tonsillarum* was identified in both Late Pleistocene woolly mammoths and the ~1.1-million-year-old Steppe mammoth (sample P033). To assess the authenticity of this microbe's ancient origin, we calculated the genetic divergence measured as mutations per base pair between the 1.1-million-year-old genome and those from the Late Pleistocene using the *PanACoTA* core-gene alignment (Figure S3A). All Late Pleistocene *E. tonsillarum* genomes exhibited high divergence from the ancient genome (0.012–0.014 mutations per base pair), whereas divergence among the Late Pleistocene genomes themselves was nearly half as high (0.006–0.0095). This supports the interpretation that the 1.1-million-year-old genome is genuinely ancient and highly divergent. To further contextualize this divergence, we used a phylogenomic alignment of 10,575 microbial species from Zhu et al. (2019)⁴¹ and calculated pairwise divergence values across 25 core genes (Figure S3B). Each of these core genes displayed a distinct evolutionary rate with the observed divergence between the ancient *E. tonsillarum* genome and the Late Pleistocene genomes falling within the range of divergence typically seen across different microbial species. This further supports the authenticity and substantial divergence of the 1.1-million-year-old *E. tonsillarum* microbial genome.

S. mutans gene content characterization

To further test the potential human origin of the two mammoth *Streptococcus* found close to the *S. mutans* clade, we examined the gene content of *S. mutans* by blasting against the NR database the 1861 genes annotated in the *S. mutans* strain (GCF_006739205.1), and identifying genes with no hits in any other *Streptococcus* genome (that could therefore be seen as *S. mutans* core genes). We therefore identified 16 genes unique to the human *S. mutans* (see Table S7 and Table S8) and none of these 16 genes was identified in the two *Streptococcus*-like mammoth sequences. Since 21.9% (404/1845) of the other genes were covered ($\geq 20\%$ of the gene covered) by mammoth reads, the probability of not sequencing any of the 16 core genes by chance is approximately 0.0188, thus suggesting that the mammoth *Streptococcus* represent a unique species distinct from the human-derived *S. mutans*.

To calculate the probability that the mammoth *Streptococcus* indeed does not contain any of the 16 *S. mutans* core genes, given that we sampled 1086 genes out of 1845 and found zero of the core genes, we used the concept of the hypergeometric distribution.

Hypergeometric Distribution

The hypergeometric distribution describes the probability of k successes (core genes in this case) in n draws from a population of size N containing exactly K successes.

In our scenario:

- $N=1845N = 1845N=1845$ (total number of *S. mutans* genes)

- $K=16$ (number of *S. mutans* core genes)
- $n=404$ (number of genes sampled drawn with a cov $\geq 20\%$)
- $k=0$ (number of *S. mutans* core genes sampled)

The probability mass function of the hypergeometric distribution is

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Supplemental figures

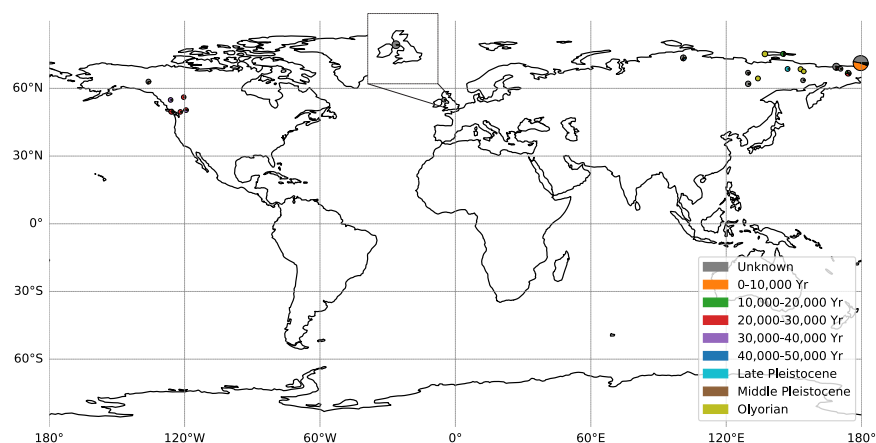


Figure S1. Distribution of mammoth samples worldwide, related to STAR Methods

The map illustrates the locations of mammoth samples used in this study, and the size of the corresponding pie charts represents the number of samples found in each area. Each pie chart displays color proportions representing the estimated age of the samples.



This figure outlines the bioinformatics workflow used to detect and analyze ancient microbial DNA from mammoth samples. The process starts with the extraction of all FASTQ sequences that did not align with *BWA-aln* v.0.7.8 (-l 16500 -n 0.01 -o 2)⁷² to either the Asian elephant, human genome, or woolly mammoth mitogenome. These unmapped reads are then classified using *Kraken2* v.2.1.2⁷⁷ against the GTDB database and further analyzed by competitive mapping against the top classified species for each genus (as identified by *Kraken2*). The pipeline includes steps to filter out contaminants, assess coverages, and evaluate the ancient nature of the reads by evaluating damage patterns using *Amber* v.1,⁹⁰ read lengths, and calculating PMD scores using *PMDtools* v.0.6.¹⁸ Simultaneously, reads from ancient sediments as well as laboratory blank reads are mapped to the same microbial references to identify putative environmental contaminants. In the next step, only known zoonotic microbes were selected for further analysis. Sequence alignment was then performed with *PanACoTA* v.1.4.1²⁹ by first aligning modern microbial gene assemblies from a given genus. Candidate ancient reads from either contamination or mammoth sources were then mapped back to the genetically closest microbial assembly in the alignment. Phylogenetic inferences were then performed using *IQ-TREE* v.2.3.5²⁸ and phylogenetic placements using *RxML EPA-ng* v.0.3.8⁸⁵ and *pathPhynder* v.1.⁸⁷ Finally, classification of putative virulence genes was performed by first mapping all reads from a monophyletic clade in the phylogeny representing a unique species and then recovering the putative ancient mosaic genome with *ANGSD* v.0.940.⁷¹ The resulting fasta file was then aligned against the virulence factor database (VFDB)⁸⁸ using *Mmseqs* search v.15-6f452.⁸⁹ The aligned sequences were translated to amino acids and used as input to a *VirulentPred* v.2.0³⁰ to detect patterns of virulent proteins.

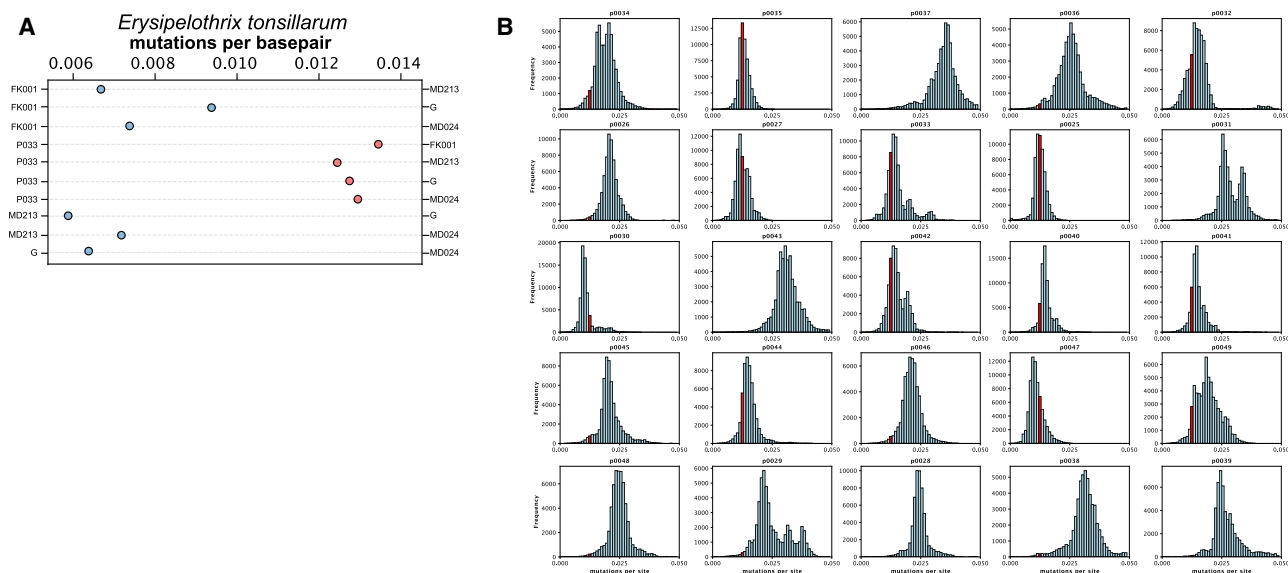


Figure S3. Sequence divergence among microbial genomes, related to STAR Methods and Figure 4

(A) Sequence divergence between *Erysipelothrix tonsillarum* genomes from mammoth samples. Divergence is based on the alignments of conserved core genes, with pairwise differences calculated by dividing the number of differing sites by the total number of sites covered by both samples. The 1.1-million-year-old sample (P033) exhibits significantly greater divergence from the Late Pleistocene samples than the divergence observed between the Late Pleistocene samples themselves.

(B) Sequence divergence for 25 core genes across a phylogeny of 10,575 microbial species, highlighting pairwise differences. The divergence of the 1.1-million-year-old sample (P033) to the other Late Pleistocene samples is shown in red. Each plot represents a single gene, and although the genes evolve at different rates, P033's divergence falls within the typical range observed for microbial species.

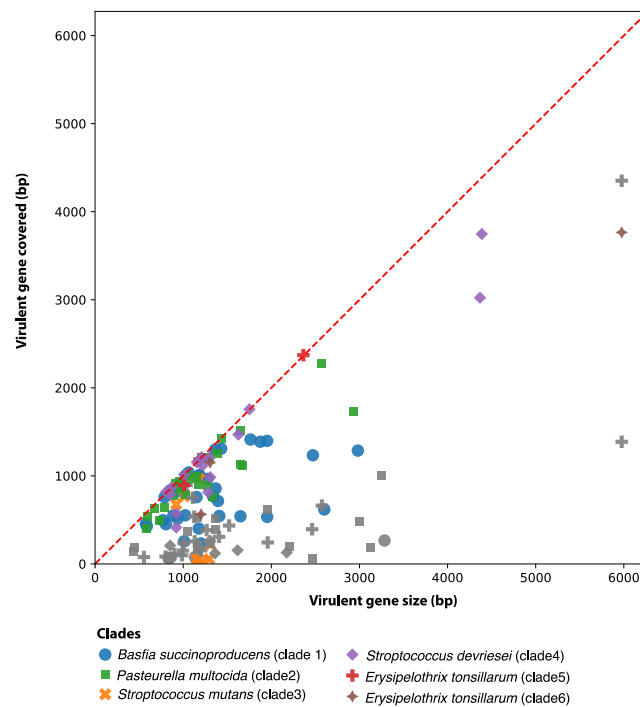


Figure S4. Scatter plot comparison of covered versus total size (in base pairs) of virulent genes, related to STAR Methods and Figure 4
Each point represents a gene, with colors and shapes indicating the reference name to which the reads have been mapped. Gray points denote genes with coverage below the coverage percentage of the reference assembly (coverage clade 1 = 11.5%, clade 2 = 50%, clade 3 = 1.8%, clade 4 = 28.5%, clade 5 = 9% and clade 6 = 88%). The dashed red line represents a perfect match where a candidate gene is totally covered, serving as a reference for alignment consistency.

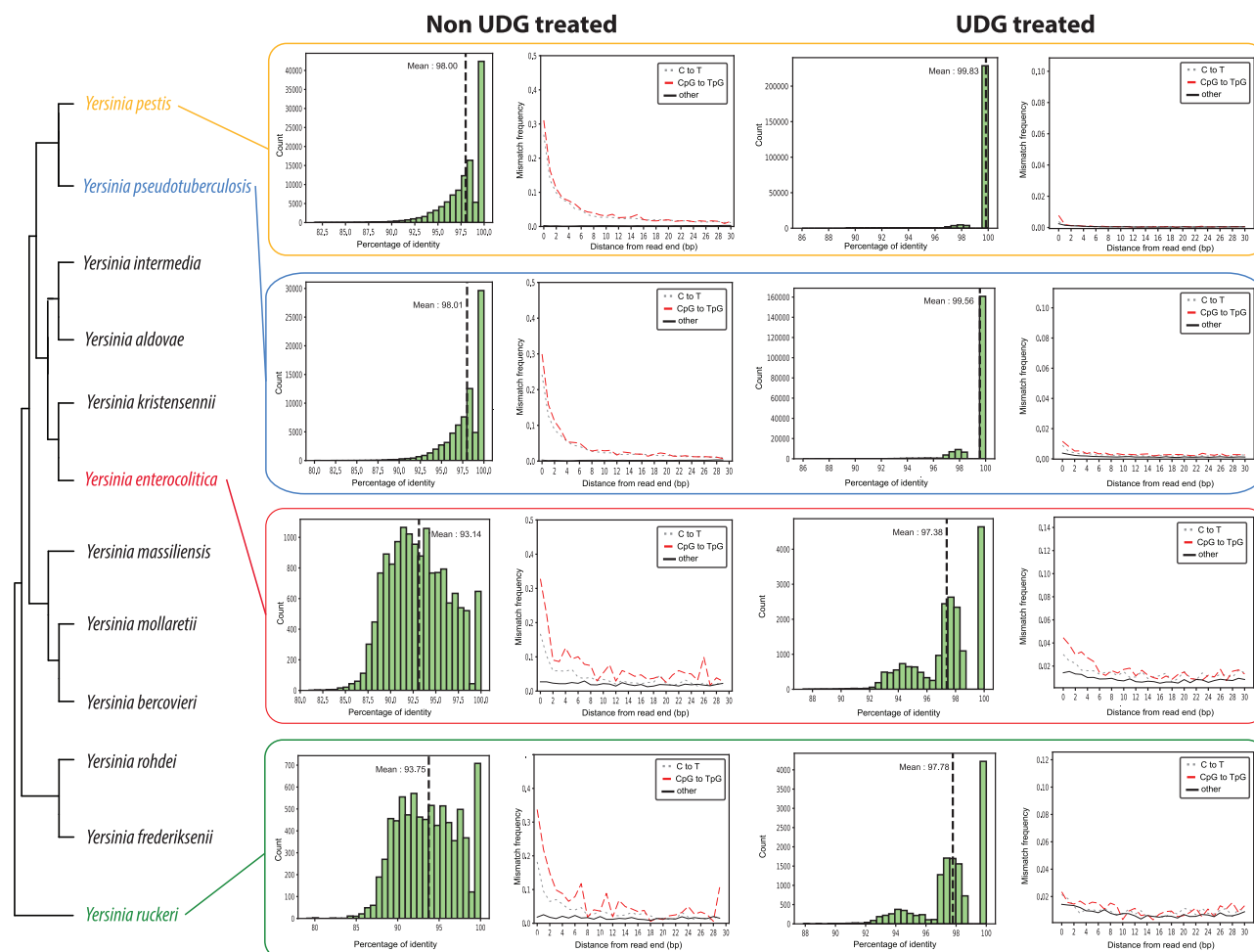


Figure S5. Damage plot distribution of *Y. pestis* reads mapped to different *Yersinia* reference genomes, related to STAR Methods and Figure 2

The left side of the figure shows the *Yersinia* phylogenetic cladogram adapted from Tan et al.⁹¹ and the colored taxa correspond to genomes to which Gok2 *Y. pestis* reads were mapped. Non-UDG-treated reads are shown on the left, and the UDG-treated reads are on the right. On the right, two images are shown for each species. The left image corresponds to the identity distribution of the mapped reads, and the right image depicts the mismatch frequency along the read sequences, where high mismatch frequency is expected at the beginning of the molecule, while less is expected in the middle part.

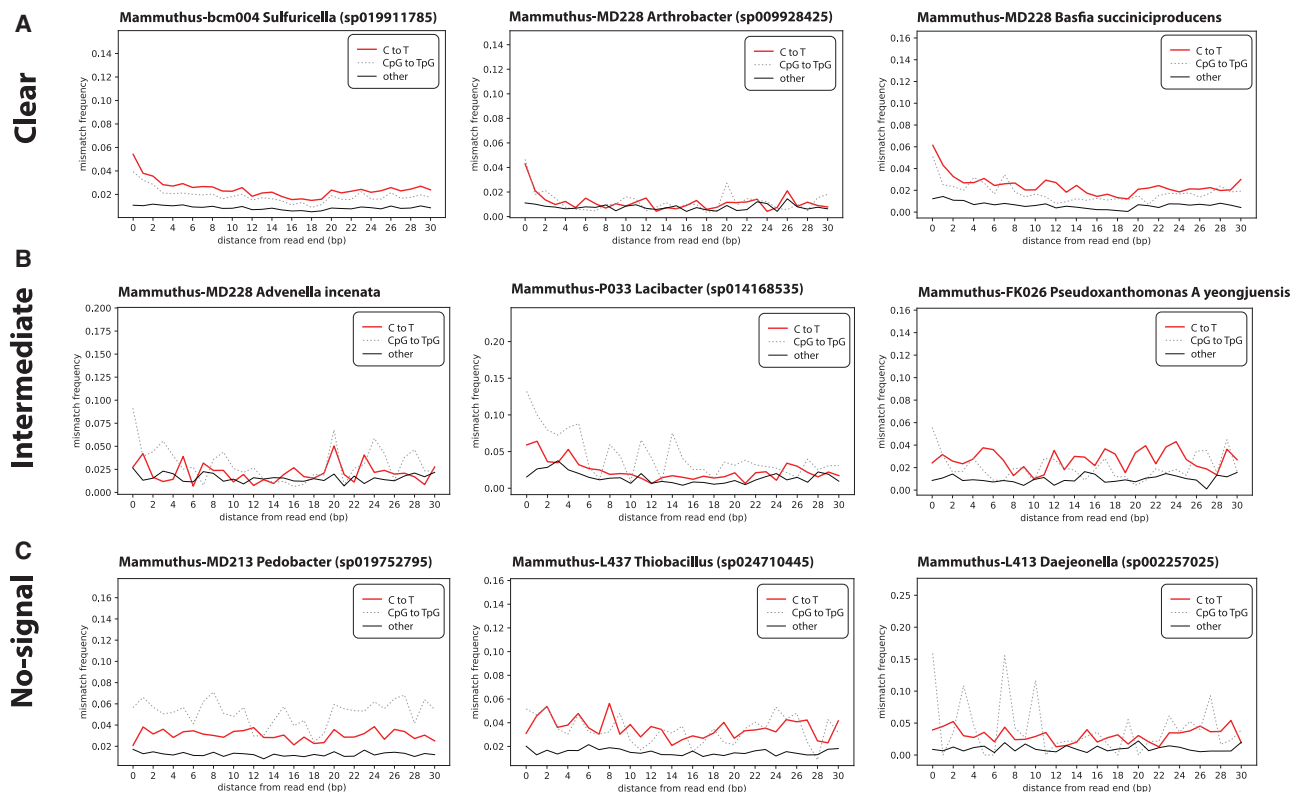


Figure S6. Damage plot classifications, related to STAR Methods

All microbes in these graphs were classified with a score ≥ 4 in our detection pipeline. We proceeded by manually examining and sorting the DNA damage patterns of these candidates into three distinct groups. First (A), those exhibiting a conspicuous CpG > TpG (or C > T for non-UDG-treated samples) trend, characterized by damage levels at the beginning of the sequence and minimal damage in the middle sections, alongside consistent patterns across all read lengths. Second (B), samples displayed damage plots featuring discrepancies in CpG > TpG and/or C > T sites, with damage decreasing from the ends toward the center of the mapped reads, albeit with some stochasticity. Last (C), cases where the damage plots displayed excessive stochasticity with no discernible pattern of decreasing damage.