# The landscape of ancient human pathogens in Eurasia from the Stone Age to historical times

Martin Sikora[1,*], Elisabetta Canteri[1], Antonio Fernandez-Guerra[1], Nikolay Oskolkov[2], Rasmus Ågren[3], Lena Hansson[4], Evan K. Irving-Pease[1], Barbara Mühlemann[5,6], Sofie Holtsmark Nielsen[7], Gabriele Scorrano[1], Morten E. Allentoft[1,8], Frederik Valeur Seersholm[1], Hannes Schroeder[1], Charleen Gaunitz[1], Jesper Stenderup[1], Lasse Vinner[1], Terry C. Jones[5,6,9], Bjørn Nystedt[10], Julian Parkhill[11], Lars Fugger[12,13], Fernando Racimo[1], Kristian Kristiansen[1,14], Astrid K. N. Iversen[11,15], Eske Willerslev[1,16,17*]

[1]Globe Institute, University of Copenhagen, Copenhagen, Denmark
[2]Department of Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Lund University, Lund, Sweden
[3]Department of Biology and Biological Engineering, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Chalmers University of Technology, Göteborg, Sweden
[4]Definitive Healthcare, Gothenburg, Sweden.
[5]Institute of Virology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, 10117 Berlin, Germany.
[6]German Centre for Infection Research (DZIF), partner site Charité, 10117 Berlin, Germany
[7]Statens Serum Institut, Copenhagen, Denmark
[8]Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, Australia
[9]Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge, UK
[10]Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[11]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK
[12]Oxford Centre for Neuroinflammation, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, UK
[13]MRC Human Immunology Unit, John Radcliffe Hospital, University of Oxford, Oxford, UK
[14]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden.
[15]Nuffield Department of Clinical Neurosciences, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[16]Department of Zoology, University of Cambridge, Cambridge, UK
[17]MARUM Center for Marine Environmental Sciences and Faculty of Geosciences, University of Bremen, Bremen, Germany

* Corresponding authors; email: martin.sikora@sund.ku.dk; ew482@cam.ac.uk

## Summary

Infectious diseases have had devastating impacts on human populations throughout history. Still, the origins and past dynamics of human pathogens remain poorly understood[1]. To create the first spatiotemporal map of diverse ancient human microorganisms and parasites, we screened shotgun sequencing data from 1,313 ancient human remains covering 35,000 years of Eurasian history for ancient DNA deriving from bacteria, viruses, and parasites. We demonstrate the widespread presence of ancient

microbial DNA in human remains, identifying over 2,400 individual species hits in 896 samples. We report a wide range of pathogens detected for the first time in ancient human remains, including the food-borne pathogens *Yersinia enterocolitica* and *Shigella* spp., the animal-borne *Leptospira interrogans*, and the malaria-causing parasite *Plasmodium vivax*. Our findings extend the spatiotemporal range of previously described ancient pathogens such as *Yersinia pestis*, the causative agent of plague, *Hepatitis B virus*, and *Borrelia recurrentis*, the cause of louse-borne relapsing fever (LBRF). For LRBF we increase the known distribution from a single medieval genome to 31 cases across Eurasia covering 5,000 years. Grouping the ancient microbial species according to their type of transmission (zoonotic, anthroponotic, sapronotic, opportunistic, and other), we find that most categories are identified throughout the entire sample period, while zoonotic pathogens, which are transmitted from living animals to humans or which have made a host jump into humans from animals in the timeframe of this study, are only detected from ~6,500 years ago. The incidence of zoonotic pathogens increased in our samples some 1,000 years later before reaching the highest detection rates ~5,000 years ago, and was associated with a human genetic ancestry component characteristic of pastoralist populations from the Eurasian Steppe. Our results provide the first direct evidence for an epidemiological transition to an increased burden of zoonotic infectious diseases following the domestication of animals[2]. However, they also reveal that the spread of these pathogens first becomes frequent thousands of years after increased animal-human contact, likely coinciding with the pastoralist migrations from the Eurasian Steppe[3,4]. This study provides the first spatiotemporal map of past human pathogens using genomic paleoepidemiology, and the first direct evidence for an epidemiological transition of increased zoonotic infectious disease burden after the onset of agriculture, through historical times.

## Introduction

Pathogens have been a constant threat to human health throughout history. Until approximately 1850, at least a quarter of all children passed away before age 1, and around another quarter before turning 15. Infectious diseases are estimated to have been responsible for over half of these deaths[5]. Not surprisingly, outbreaks have had profound impacts on human societies, at times devastating entire civilizations[6,7]. The burden of infectious diseases has also left lasting impressions on human genomes, where the selective pressure exerted by pathogens has continuously shaped human genetic variation[8,9]. Where and when different human pathogens first emerged, how and why they spread, and how they affected human populations are important but largely unresolved questions for human health.

During the Holocene (beginning ~12,000 years ago), the agricultural transition resulted in more sedentary and larger communities, facilitating pathogen transmission within populations[10]. Simultaneously, the rise of animal husbandry and pastoralism increased the risk of zoonoses. Additionally, technological advances, such as the use of horses and carts for increased mobility, heightened the risk of disease transmission between populations[11]. It has been hypothesised that these changes created the conditions for a so-called "first epidemiological transition" characterised by increased infectious disease mortality[2]. However, direct evidence for such a transition remains scarce[12]. Paleopathological examinations of ancient skeletons have provided some insights into past infectious disease burden[13], but are limited to the small number of infections that leave detectable marks on bone tissue. More recently, advances in ancient DNA techniques have provided the means to retrieve direct genomic evidence of past microbial infections, even in the absence of visible disease pathologies, and in some cases allows for the

reconstruction of complete ancient pathogen genomes. These studies have often provided surprising insights into the evolutionary history, and genome evolution, of the causative agents of some of the most important infectious diseases affecting humans, including plague (*Yersinia pestis*)[11,14–22], tuberculosis (*Mycobacterium tuberculosis*)[23,24], smallpox (*Variola virus*)[25,26], Hepatitis B (*Hepatitis B virus*)[27–29] and others[30–35]. Nevertheless, these studies have focussed on the evolution of individual pathogens, where whole genomes could be recovered. A comprehensive investigation of the broader landscape of ancient pathogens across multiple microorganisms, geographic regions and time periods, to characterize the burden of pathogens affecting our ancestors, has been lacking.

To overcome these limitations, we developed an accurate and scalable workflow to screen for ancient microbial DNA and applied it to 1,313 ancient Eurasian individuals with shotgun-sequenced metagenomic data. We use these data to investigate long-standing questions in paleoepidemiology, such as the emergence and spread of major human pathogens and the factors influencing their dynamics.

## Workflow and dataset overview

To identify ancient microbial DNA in large-scale ancient human metagenomic shotgun sequencing datasets, we developed a high-throughput computational workflow (Fig. 1a), which consists of three main components. First, sequencing reads are subjected to an efficient *k*-mer-based metagenomic classification using *KrakenUniq*[36] against a comprehensive database of complete bacterial, archaeal, viral, and protozoan genomes from the RefSeq database[37] and the human reference genome. This allows rapid grouping of the complete set of sequencing reads according to taxonomic rank, with no requirement for computationally more costly sequence alignment. We use the taxonomic rank of genus as the unit of grouping. Second, the set of reads for a particular genus are aligned with *minimap2*[38] against each best-matching reference assembly for all species within the respective genus, selected based on the maximum number of unique *k*-mers assigned to the assembly by *KrakenUniq*. Finally, summary statistics quantifying the similarity of the reads to the reference assembly, ancient DNA damage patterns, and evenness of genomic coverage are calculated from the resulting species-level alignments to confirm true positive ancient microbial hits (Fig. 1a; Methods).

We applied this workflow to a dataset of ~405 billion ancient DNA sequencing reads from 1,313 ancient Eurasian individuals, spanning a ~37,000 year time period from the Upper Paleolithic to historical times (Fig. 1b; Supplementary table S1). In the initial taxonomic classification, the fraction of reads classified as originating from non-human taxa ranged from 1.1% to 76.5% per sample, with a median of 3.5%. The genera with the highest fraction of reads classified include soil-dwelling taxa such as *Streptomyces* and *Bradyrhizobium* or taxa with broad ecological distribution, such as *Pseudomonas*, reflecting a predominantly environmental source of microbial DNA in the samples (Extended Data Fig. 1). To further characterise sample metagenomic composition, we applied topic model analysis. In this analysis, a sample is modelled as a mixture of *k* discrete "topics", each representing a component of the metagenome determined by distinct patterns of relative abundances of genus-level *k*-mer counts. Assuming k=3 components, we found that the two components with the highest proportions across all samples were driven by the highly abundant environmental taxa *Streptomyces* / *Bradyrhizobium* or *Pseudomonas* (Extended Data Fig. 1d). Interestingly, the third component was driven by the highest relative abundance of genera such as *Actinomyces*, *Streptococcus* or *Pseudopropionibacterium*, commonly associated with the human oral microbiome. This component was observed at noticeable frequencies in many samples

derived from teeth, but only rarely in those from other tissues (Extended Data Fig. 1e, f ). This suggests that a substantial fraction of the identifiable microbial DNA found in many ancient tooth samples derives from the endogenous oral microbiome of the human host.

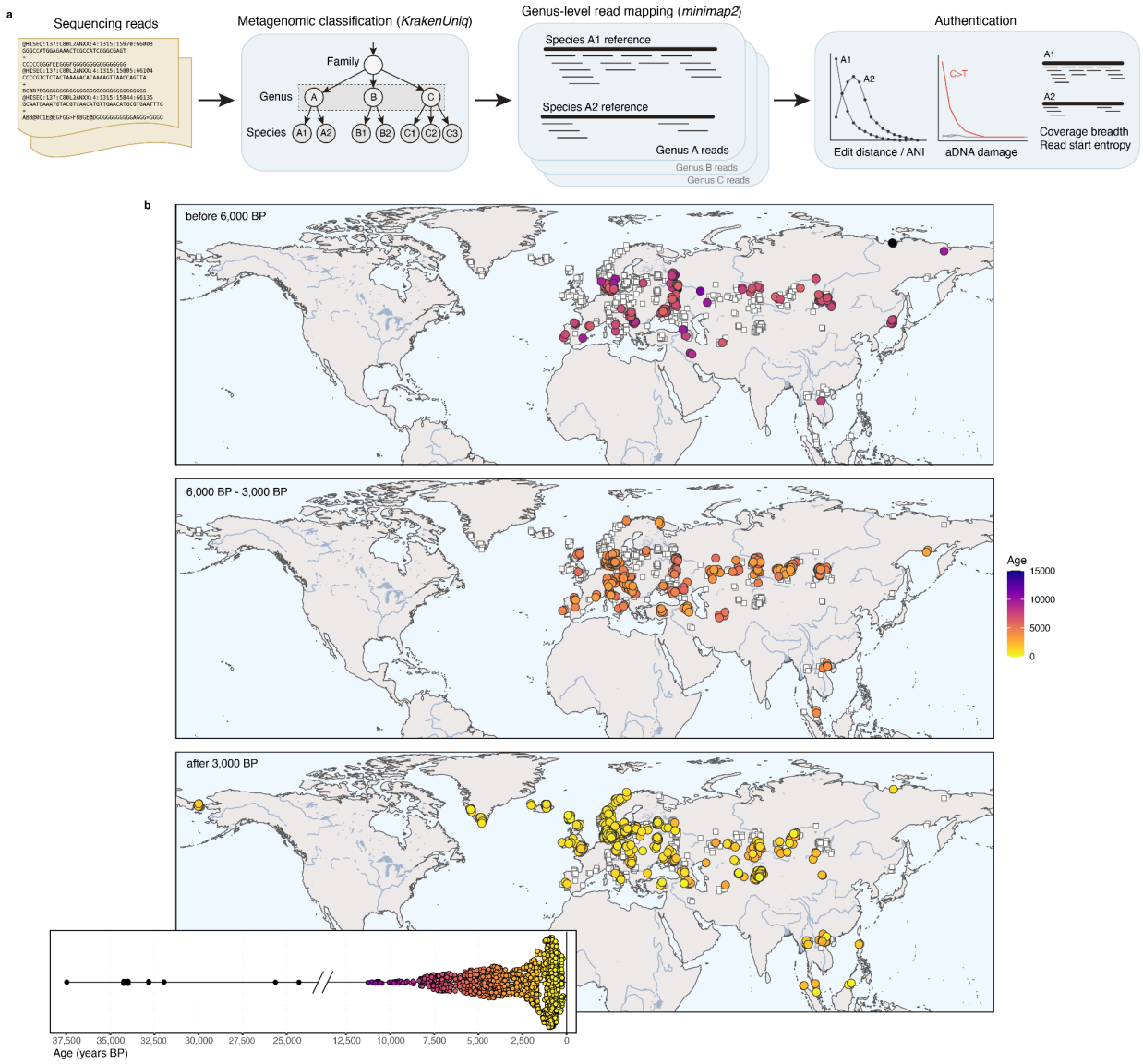



**Fig. 1. Ancient pathogen screening workflow and dataset. a,** Overview of the main components of the ancient pathogen screening workflow. **b,** Spatiotemporal distribution of the study samples. White squares indicate locations of the full set of 1,313 study samples, whereas coloured circles highlight location and age of samples from the time period indicated in the respective panel.

## Ancient microbial DNA in remains of 1,313 Eurasians

We selected a set of 44 bacterial and protozoan genera containing known human pathogens to screen for ancient microbial DNA in the genus-level read mapping and authentication stages (Extended Data Fig. 2, 3). To increase the resolution for identifying ancient viral DNA, we reanalysed all non-human classified

reads with a viral-specific database using the NCBI full viral neighbour assemblies dataset of validated complete or nearly complete viral genomes in GenBank, covering 1,356 genera. For each genus of interest in each sample, we selected the species passing all authentication criteria (Methods) and with the highest number of unique *k*-mers assigned by *KrakenUniq* (Extended Data Fig. 2a,b) as the putative ancient microbial species hit. We found that ancient microbial DNA was widely detected, with 2,474 individual hits identified across 896 samples (Supplementary table S2). We further distinguished these putative ancient microbial hits by the average nucleotide identity (ANI) of the mapped sequencing reads to the reference assembly (high or low, ANI cutoff 97%) and by the evidence for ancient DNA damage using *metaDMG* (lenient, Z-score ≥ 0.5; strict, Z-score ≥ 2). When restricting to the most stringent group (high ANI / strict metaDMG), 594 ancient microbial hits across 372 samples remained (Fig. 2a; Extended Data Fig. 4, 5). Across all ancient microbial hits, the highest number were observed in bacterial genera commonly found in soil environments, such as *Clostridium* (360; 27% of all samples) and *Mycobacterium* (188; 14% of all samples), or genera associated with the human microbiome, such as *Streptococcus* (302; 23% of samples) and *Treponema* (237; 18% of samples) (Fig. 2a; Extended Data Fig. 4, 5). The relative abundance of ancient microbial reads varied by multiple orders of magnitude, from highly abundant hits in species such as Pseudomonas stutzeri (up to 7% of all non-human classified reads mapped; median 0.06%) to hits at the lower limits of detection as, e.g., for the louse-borne pathogen Borrelia recurrentis (lowest abundance ~1 read mapped per 1,000,000 non-human classified; median 0.0026%; Extended Data Fig. 6a). Ancient microbial DNA from species commonly found in soil such as *Clostridium botulinum* was detected at similar rates in tooth and bone samples. Conversely, species associated with the human oral microbiome (e.g. *Corynebacterium matruchotti, Treponema denticola, and Streptococcus sanguinis*) or pathogenic infections (e.g. *Yersinia pestis* and *Hepatitis B virus*) were significantly more frequently identified in tooth samples (Fig. 2b).

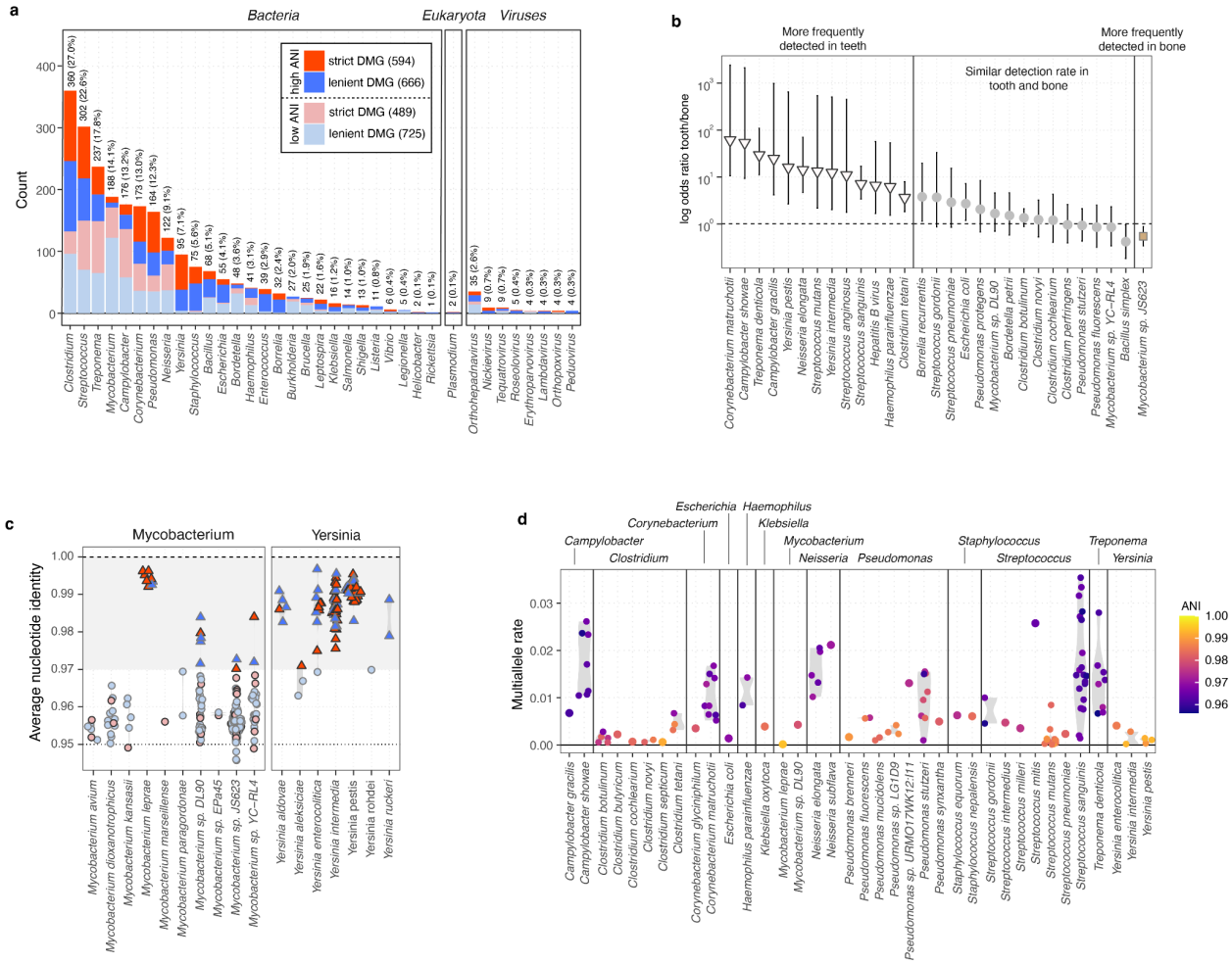



**Fig. 2. Overview and characteristics of detected ancient microbial DNA. a,** Barplot showing total number of putative ancient microbial hits (overall detection rate in brackets) for bacterial, eukaryotic and viral ($n \geq 4$) genera. Bar colour and shading distinguishes counts in the different ANI and aDNA damage categories. **b,** Odds ratios for association of ancient hits with sample material (tooth or bone) across 29 species with $\geq 20$ ancient hits. Symbols indicate significance of association ($p \leq 0.01$, Fisher's exact test; white triangles - more frequently identified in tooth; brown squares - more frequently identified in bone; grey circles - no significant association). Error bars indicate 95% confidence interval of odds ratio. **c,** Distributions of ANI for individual species hits in the genera *Mycobacterium* and *Yersinia*. Symbol colour and shape indicate ANI (triangle - high ANI; circle - low ANI) and aDNA damage category (red - strict metaDMG; blue - lenient metaDMG). Grey shading indicates ANI range for high ANI hits (ANI $\geq 0.97$) **d,** Rates of observing multiple alleles at genomic sites (read depth $\geq$ 2X) in 115 ancient hits (average read depth $\geq$ 1X) across 37 species. Symbol colour indicates ANI.

We observed marked differences in the distributions of the genetic similarity of the ancient microbial sequences (measured using ANI) to their reference assemblies, both among genera and between species within a genus (Fig. 2c; Extended Data Fig. 6). High ANI can be indicative of ancient microbial sequences closely related to a reference assembly in the database and was frequently observed in hits across all species from genera such as Yersinia and Staphylococcus. In other genera, only few hits had a closely related reference assembly match in the database. One such example is the genus Mycobacterium, where only hits of the leprosy-causing bacterium Mycobacterium leprae were highly similar to their reference assembly (ANI > 99%; Fig. 2a,c). Low ANI indicates that the ancient microbial DNA is only distantly related to the reference assembly used for mapping, for example, due to poor representation of the diversity of the genus in the database or false-positive classification of ancient microbial reads

deriving from a related genus within the same family. Alternatively, ANI can also be reduced when reads mapped to a particular reference assembly are derived from multiple closely related genomes present in a sample. To test for the presence of multiple closely related genomes, we quantified the rate of observing multiple alleles at nucleotide positions across the genomes of hits with read depth ≥1X (Fig 2d). We found an enrichment of multiple alleles in several species linked to the human oral microbiome, such as *Campylobacter showae* associated with gingivitis and periodontitis[39], as well as some environmental species, including *Clostridium tetani*. Hits for these species were also associated with low ANI, consistent with the expectation for mixtures of ancient microbial DNA (Fig. 2d).

Overall, our results show that ancient microbial DNA isolated from human remains often originates from complex mixtures of distinct endogenous and exogenous sources. The high detection rate, high relative abundance, lower ANI, and evidence of multiple genome mixtures in genera such as *Clostridium*, *Pseudomonas* or *Mycobacterium* suggest that a substantial fraction of the ancient microbial metagenome associated with human remains derives from environmental sources, possibly involved in post-mortem putrefaction processes. By contrast, species from other frequently observed genera, including *Streptococcus, Campylobacter* or *Neisseiria,* were predominantly identified from teeth and likely originated from the endogenous oral microbiome of the individual. Species representing likely cases of actual infections with known human pathogens (e.g. *Yersinia pestis* and *Mycobacterium leprae*) were often characterised by higher ANI and/or low multi-allele rate, consistent with pathogen load originating predominantly from a single dominant strain.

## The landscape of ancient pathogens across Eurasia

Our ancient pathogen dataset provides a unique opportunity to investigate the origins and spatiotemporal distribution of major human pathogens in Eurasia. Analyzing these data allowed us to expand the spatiotemporal range of many previously identified ancient pathogenic species, as well as reporting species identified for the first time using paleogenomic data.

Considering bacterial pathogens, we found widespread distribution of the plague-causing bacterium *Yersinia pestis*, consistent with previous studies[11,15,16,18,21,40,41]. We identified 39 cases of *Y. pestis*, corresponding to a detection rate in ~3% of our samples. These newly identified cases expand both the spatial and temporal extent of ancient plague over previous results (Fig. 3). The earliest three cases were dated between approximately 5,700-5,300 cal. BP, across a broad geographic area ranging from Western Russia (NEO168, 5,600-5,300 cal. BP), Central Asia (BOT2016, 5,600-5,300 cal. BP, Botai culture) to Lake Baikal in Siberia (5,700-5,400 cal. BP). This broad range of detection among individuals pre-dating 5,000 cal. BP challenges previous interpretations that early plague strains represent only isolated zoonotic spillovers[19]. In Europe, we report several new early cases, including an individual from the Iberian peninsula (NEO650, 4,200-4,000 cal. BP), and the first evidence of ancient plague from the far northwestern limits of the continent, in an individual from a Neolithic burial site on the island of Orkney in Scotland (Banks chambered tomb; NEO630, 4900-4800 cal. BP), pre-dating the previously reported first arrival of plague in Britain by ~900 years[42]. We replicated previously identified cases of plague in Late Neolithic and Bronze Age (LNBA) contexts across the Eurasian Steppe[15] and identified new cases from the same burial context in multiple sites (RISE509, RISE511, newly identified RISE510, Afanasievo Gora, Russia; RISE505, newly identified RISE500, Kytmanovo, Russia; RISE397, newly identified RISE396, Kapan, Armenia). These results indicate that the transmissibility and potential for

local epidemic outbreaks for strains at those sites were likely higher than previously assumed[19]. Finally, 11 out of 39 cases were identified in individuals from three medieval cemeteries in Denmark, highlighting the high burden of plague during this time in Europe.

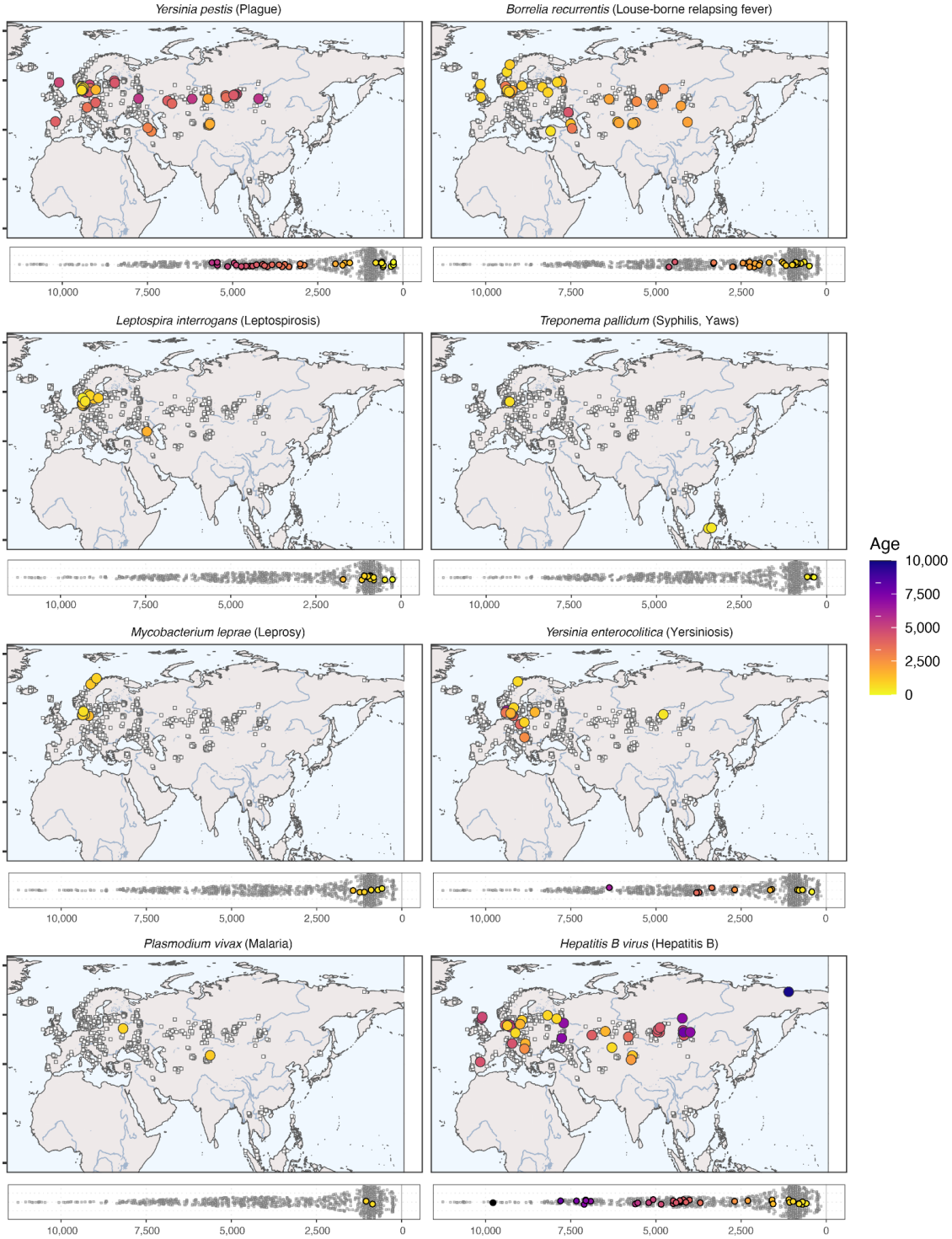



**Fig. 3. Spatiotemporal distribution of selected ancient pathogens.** Each panel shows geographic distribution (top) and timeline (bottom) for identified cases of the respective pathogen (indicated by coloured circle). Geographic locations and age distributions of all 1,313 study samples are shown in each panel using white squares.

Another bacterial pathogen frequently detected was the spirochaete bacterium *Borrelia recurrentis*, the causative agent of louse-borne relapsing fever (LBRF), an epidemic disease with a mortality of 30-70%[43]. While previous paleogenomic evidence for LBRF is limited to a single genome from 15th-century Scandinavia[30], we now report 31 cases (2.3% detection rate), within a wide geographic region encompassing Europe, Central Asia and Siberia (Fig. 3). We detected the earliest cases in LNBA individuals from Scandinavia (RISE61, 4,800-4,400 cal. BP; RISE94, 4,600-4,400 cal. BP) but found the overall highest detection rates during later time periods (Iron and Viking Ages). Outbreaks of LBRF were historically associated with poor living conditions and personal hygiene but are rare today[43]. Our results thus suggest that this disease exerted a substantially higher burden on past populations. We also detected other bacterial pathogens with previous evidence from paleogenomic data. The leprosy-causing bacterium *Mycobacterium leprae* was identified in eight individuals (0.6% detection rate), all of them from Scandinavia and only appearing from the Late Iron Age onwards (earliest case RISE174, 1,500-1,300 cal. BP). These findings are consistent with widespread distribution of leprosy in medieval Europe, as previously suggested[44]. We further detected three cases of *Treponema pallidum*, subspecies of which are the causative agents of treponematoses such as yaws, as well as endemic and venereal syphilis, in three individuals from more recent time periods (earliest case 101809T, Denmark, 600-500 BP; Fig. 3). Two cases were identified in individuals from the island of Borneo in Southeast Asia (approximately 500-300 years BP), to our knowledge the first paleogenomic evidence for treponemal disease from the region.

Our broad screening allows us to report multiple instances of ancient bacterial pathogen species with the first direct molecular evidence from paleogenomic data. Within the genus of *Yersinia,* we identified twelve cases of *Yersinia enterocolitica,* the causative agent of yersiniosis, commonly contracted through consuming contaminated raw or undercooked meat. The animal reservoirs for *Y. enterocolitica* include horses, cattle, and sheep. As *Y. enterocolitica* rarely enters the bloodstream, our results likely underestimate the disease burden. We also detected other members of the order *Enterobacterales* transmitted via the fecal-oral route, including members of the genera *Shigella*, *Salmonella*, and *Escherichia*. As species within those genera are genetically closely related, we further investigated their co-occurrence within individuals across the dataset. Out of 82 identified putative hits (the majority of which were *Escherichia coli*; 55 cases, 4.1% of samples), 27 were found as co-occurring hits in a set of 12 individuals (Extended Data Fig. 7). Restricting these putative cases to the species with the highest ANI in the individual, and removing low ANI hits and *E. coli* hits due to their ubiquitous distribution, we were left with 6 cases of *Shigella* (*Shigella dysenteriae* n=3*; Shigella boydii* n=2; and *Shigella flexneri* n=1) and 5 cases of *Salmonella enterica*, previously identified in ancient Europeans from Neolithic contexts[34]. Interestingly, these cases of putative food-borne infections include some of the only ancient pathogens we identified in individuals from Mesolithic hunter-gatherer contexts (NEO941, Denmark, 6,400-6,300 cal. BP - *Yersinia enterocolitica*; NEO19, Denmark, 8,300-8,000 cal. BP - *Shigella dysenteriae*).

Among diseases where exposure to infected animals constitutes an important risk factor, we report the first evidence for ancient leptospirosis (genus *Leptospira*), in the earliest case dating back to 5,600-5,400 cal. BP (NEO46, Sweden). Most of the identified hits were found to be *Leptospira interrogans* (n=15, 1.1% detection rate), almost exclusively in Scandinavian contexts from the Viking Age onwards (Fig. 3). In contrast, the earlier cases involved predominantly *Leptospira borgpetersenii* (n=4, 0.3% detection rate), which today primarily is found in cattle while *L. interrogans* is seen more broadly in wild and

domesticated animals. Notably, none of the cases of leptospirosis involved samples from hunter-gatherer contexts. Other diseases known to occur in association with animals and livestock could not be reliably identified, either because of weak evidence for ancient DNA damage (e.g. listeriosis, *Listeria monocytogenes*), or putative hits with low ANI deriving from non-pathogenic species (e.g. brucellosis; genus *Brucella*). Examples of other major human pathogens not identified in our dataset include *Mycobacterium tuberculosis* and *Vibrio cholerae*. However, both species rarely cause bacteremia, which is usually restricted to immunocompromised patients, and are hence unlikely to be identified using metagenomic data from tooth and bone remains sampled for ancient human DNA.

Identifying eukaryotic pathogens is challenging, as sequence contamination from other organisms regularly occurs in reference genomes deposited in public databases[45]. Illustrative examples in our dataset are protozoan parasites such as *Toxoplasma gondii*, which were readily identified as hits with high ANI and ancient DNA damage, but with low support from coverage evenness statistics due to the ancient reads predominantly mapping to short contigs representing contamination (Extended Data Fig. 3). Despite these challenges, we identified two cases of infections with the malaria-causing parasite Plasmodium vivax, one in a medieval individual from Central Asia (DA204, Kazakhstan; 1,100-1,000 cal. BP) and one in a Viking Age individual from Eastern Europe (VK224, Russia; 950-750 BP; Fig. 3). To our knowledge, these constitute the first ancient metagenomic evidence for malaria from archaeological remains. The *P. vivax* malaria vector *Anopheles atroparvus* might in the past, like today, have been widespread in Europe and surrounding areas, including, e.g., North-Eastern Europe, Russia, and the Pontic Steppe[46,47]. Our results directly demonstrate the occurrence of *P. vivax* malaria at northern latitudes ~1,000 years ago and indicate metagenomic datasets from ancient human remains as viable means to study the past structure of parasite populations in Europe prior to their elimination in the twentieth century[33,46,48].

Among viral species, we found widespread evidence for infections with *Hepatitis B virus* (HBV; 35 cases, 2.6% detection rate), consistent with previous studies[27–29]. Our newly reported HBV cases include individuals from Mesolithic (Kolyma River, n=1) and Neolithic (Lake Baikal, n=5) contexts in Siberia dating back as early as 9,900-9,600 cal. BP, providing the first evidence for ancient HBV from those regions (Fig. 3). We also report the first ancient cases (n=2) of *Torque teno virus* (TTV), small single-stranded DNA viruses, dating back ~7,000 years (NEO498, Ukraine; 7,200-6,900 cal. BP). Other ancient virus hits included viruses not known to infect humans, such as examples of ancient phage DNA (e.g. *Escherichia phage T4, Proteus virus Isfahan*) and surprisingly, one case of an ancient insect virus (*Invertebrate iridescent virus 31 (IIV-31))* in a tooth sample of a Viking Age individual from Sweden (VK30, Varnhem; 950-650 BP). The natural host for this virus is the woodlouse (oniscideans), where infection causes a characteristic change to a bluish iridescent colour.

Co-infections with multiple pathogens can exacerbate the progression and outcomes of disease in an infected individual[49] and was likely an important morbidity factor in ancient human populations. To test for co-infections in our dataset, we searched for individuals showing co-occurrence of distinct ancient microbial species likely representing true ancient infections as described above. The strongest case for an ancient co-infection was observed in a Viking Age individual from northern Norway (VK388). We replicated previous results for infection with ancient variola virus (aVARV) in this individual[26] and additionally found evidence for infection with the leprosy-causing bacterium *Mycobacterium leprae*. Another case of possible co-infection with *Mycobacterium leprae* was found in VK366, a Viking Age

individual from Denmark, showing additional evidence for leptospirosis (*Leptospira interrogans*). Among the many cases of early plague identified in the dataset, we only found a single instance of possible co-infection, a Bronze Age individual from Russia (RISE386) showing evidence of an ancient *Salmonella enterica* infection. An intriguing early case of a possible co-infection was found in a mesolithic hunter-gatherer from Russia (Sidelkino, 11300-11100 cal. BP). This individual showed evidence of two important human pathogens, the respiratory pathogen *Corynebacterium diphtheriae* causing diphtheria, and *Helicobacter pylori*, usually restricted to gastric infections; however, rare contemporary examples of bacteremia have been reported for both[50,51]. Interestingly, no other cases for these two bacteria were found across the entire dataset, suggesting a rare case of possibly severe systemic infections where both species entered the bloodstream. Overall, our results show that co-infections can be readily detected using ancient metagenomic screening.

## Temporal dynamics and drivers of epidemic pathogens

Understanding the factors affecting the dynamics of past epidemics is a major aim of paleoepidemiology. Our dataset allows us, for the first time, to address this question using direct molecular evidence for major epidemic pathogens across prehistory. We first analyzed the temporal dynamics of two epidemic pathogens with high detection rates in our dataset, *Yersinia pestis* (plague) and *Borrelia recurrentis* (Louse-borne relapsing fever - LBRF). To investigate changes in pathogen incidence over time, we performed Bayesian change-point detection and time series decomposition[52], using the detection rate of the respective pathogen in windows of 20 temporally consecutive samples as a proxy for its incidence. For ancient plague, we inferred a gradual rise in detection rate starting from ~6,000 BP, about 1,000 years after the estimated time to the most recent common ancestor of currently known ancient strains (7,100 cal. BP)[19]. It reached a first peak around ~5,000 BP, coinciding with the emergence and early spread of the LNBA- strains, which had limited flea-borne transmissibility[15,16,21], across Europe and the Eurasian Steppe (Fig. 4). Detection remained high with additional peaks for a ~3,000-year period, until an abrupt change ~2,800 BP led to a ~800 year period where no plague was observed. Starting at ~2,000 BP, plague reappeared in our dataset, culminating in the period around the first historically documented plague pandemic (~1500-1300 BP; Fig. 4). Another hiatus of ~800 years led to a rise and peak associated with the second plague pandemic ~600 BP (Fig. 4). This pattern of change coincides with the extinction of the LNBA- strains ~2,700 BP[21], and the second *Y. pestis* diversification event starting ~3,700 BP which gave rise to an extinct Bronze Age lineage (RT5, LNBA+)[17] and present-day lineages; these had increased flea-mediated transmission mechanisms favouring bubonic plague and led to all known later plague pandemics[53].

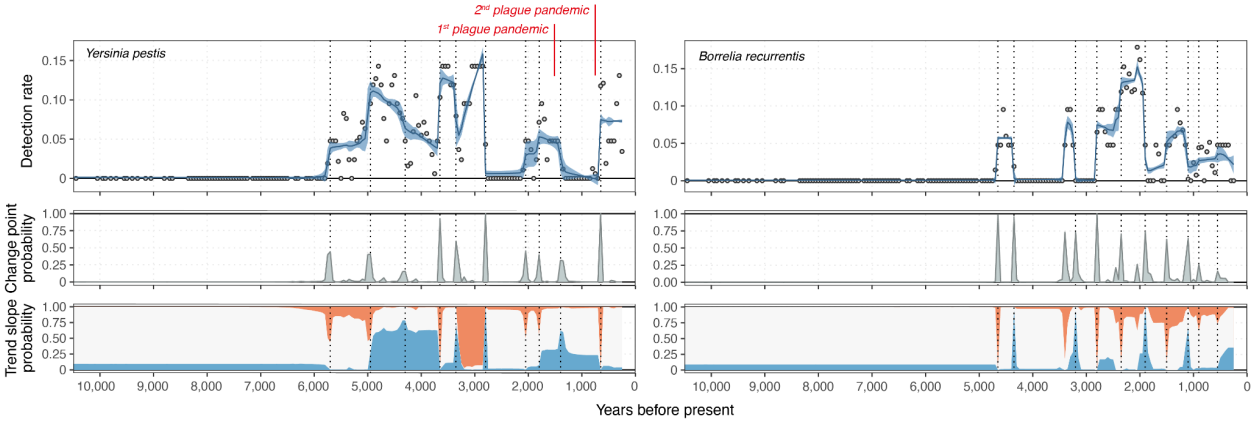



**Fig. 4. Bayesian time-series decomposition of major epidemic pathogens.** Panels show estimated trendlines and 95% credible interval for detection rates (top), probability distributions and locations (dotted lines) for change points (middle) and probability of trend slope (bottom) being positive (red), negative (blue) or zero (white), inferred using Bayesian change-point detection and time series decomposition.

The inferred temporal dynamics of LBRF show a first peak in detection around 5,500 BP, slightly more recent than for plague, but with overall more sporadic occurrences in sharper peaks during the first ~1,700 years(Fig. 4). Starting around 2,800 BP, LBRF was detected more consistently, leading to a peak around 2,000 years ago. Strikingly, this period of high LBRF detection coincided with a period without detectable plague activity (Fig. 4), suggesting that the absence of plague represents a true reduction in incidence during this period rather than sample size limitations or poor DNA preservation. LBRF remained detectable until the end of the time series, indicating the continued occurrence of LBRF epidemics across Eurasia until the medieval period.

A striking feature shared in the temporal dynamics of both plague and LBRF was the absence of detectable cases before ~6,000 BP, a time that coincides with a transition of individuals in predominantly hunter-gatherer contexts to those in farming or pastoralist cultural contexts in our dataset. It has been hypothesised that the transition from hunting and gathering to agriculture and gradually increased animal husbandry over time led to increased population densities and more frequent exposure to zoonotic diseases, facilitating the transmission of old and new pathogens[2]. Our dataset allows us to test this hypothesis using molecular evidence for infectious disease load for the first time. To increase power to detect changes in the load of different pathogen types, we first combined our ancient microbial hits into five distinct classes (Supplementary Table S3)[54]: "zoonotic" species such as *Y. pestis* which can be transmitted from living animals to humans either via direct contact, vector species or alimentary sources (n=120 hits, n=117 samples); "anthroponotic" species directly transmitted between humans without another species intermediate, and sometimes also from humans to animals (e.g. *Hepatitis B virus* or *Neisseria gonorrhoeae*) (n=71 hits, n=67 samples); "sapronotic" species transmitted from and being able to replicate in environmental sources such as soil, water, and decaying organic material (e.g. *Clostridium botulinum*) (n=343 hits, n=337 samples); "opportunistic" species commonly found as commensals in human microbiomes and only occasionally pathogenic (e.g. *Streptococcus pneumoniae*) (n=402 hits, n=308 samples); and "other" for species not covered by the previously listed classes (n=1,509 hits, n=699 samples). We note that the category "zoonotic" in this definition includes both species with frequent spill-overs from domestic or wild animals such as *Yersinia pestis* or *Listeria monocytogenes*, and vector-borne

species which have made a single host jump into humans from animals and are now human-restricted such as *Borrelia recurrenti*s or *Plasmodium vivax*[54].

Using the detection rate in consecutively sampled windows as described above to measure incidence, we found that species in the class "other" showed the highest relative detection rate across time, accounting for > 50% of ancient hits throughout most time periods (Fig. 5a). Species in the "opportunistic" class were also detected at consistent rates throughout time, accounting for ~20% of ancient hits. Species in the "sapronotic" and "anthroponotic" classes also occurred throughout all periods at lower detection rates (mostly below 10%). Strikingly, the only exception to this pattern were species in the "zoonotic" class, which were not detected before ~ 6,500 BP (Fig. 5a; Extended Data Fig. 8). Using Bayesian time series decomposition[52] on the "zoonotic" class, we inferred an overall increase in their detection rates from ~6,000 BP, thereafter remaining at similarly elevated levels until the medieval period (Fig. 5b). Our results thus provide the first direct evidence for an epidemiological transition of increased zoonotic infectious disease burden after the onset of agriculture, through historical times.

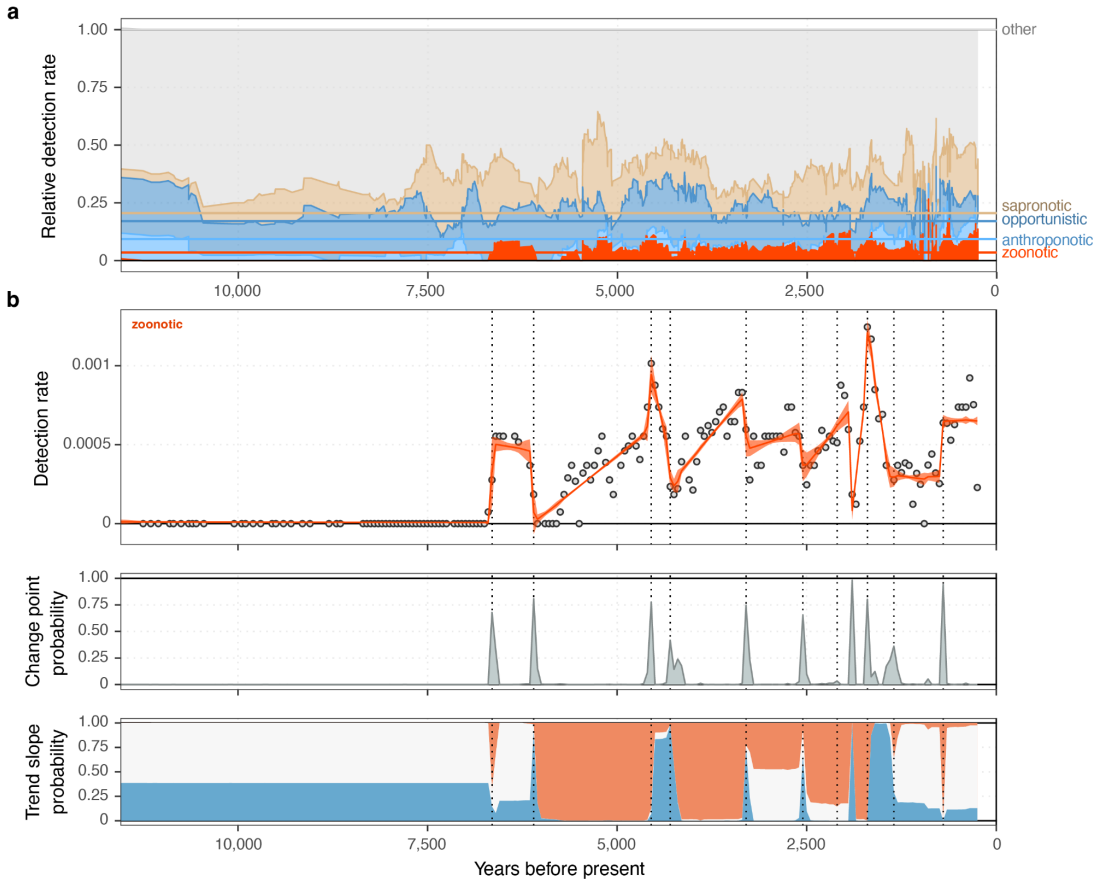



**Fig. 5. Time series of ancient microbes by transmission type. a,** Timeline of relative detection rates in sliding windows of 21 temporally consecutive samples, for different ancient microbial species classes. Coloured horizontal lines indicate the expected rates if species in all classes would be detected at equal rates, based on the total number of distinct species in each class. **b,** Trendlines for detection rates (top), probability distributions for changepoint occurrence (middle) and probability of slope for trend (bottom) inferred using Bayesian change-point detection and time series decomposition, for ancient microbial species in the "zoonotic" class.

We used Bayesian spatiotemporal modeling[55] to investigate possible drivers of the observed ancient microbial incidence. We modelled the presence/absence of either individual microbial species or combined species groups using sets of putative covariates, including spatiotemporal variables (longitude, latitude, and sample age), paleoclimatic variables (mean annual temperature and precipitation), human mobility and ancestry, as well as a proxy for "detectability" (number of human-classified reads). In the set of models for the "zoonotic" species class, the factors consistently associated with their presence/absence were the age of a sample as well as ancestry of the human host (Fig. 6; Extended Data Fig. 8). Sample age consistently was negatively associated with "zoonotic" incidence, with particularly strong effects in the individual species models for *Borrelia recurrentis* and *Leptospira interrogans*, and a smaller effect in the combined model. Ancestry of the human host, on the other hand, showed a consistent positive effect, with the Caucasus hunter-gatherer (CHG) component as the main driver. The model for *Yersinia pestis* additionally showed a positive effect of Eastern hunter-gatherer (EHG) ancestry, which in combination with CHG is a signature of so-called "Steppe" ancestry of pastoralist groups which expanded westwards and eastwards across the Eurasian Steppe during the late Neolithic and early Bronze Age. The increased incidence of zoonotic diseases in people with CHG/Steppe ancestry could reflect either an increased genetic susceptibility or the higher risk of acquiring a zoonotic disease associated with the pastoralist lifestyle. The latter suggestion seems more likely as continued exposure to selective pressure from zoonotic diseases would be expected to select for reduced susceptibility in these populations. Other types of ancestry, such as Anatolian-farmer related (LVN) or Western hunter-gatherer (WHG), human mobility or climate-related variables had no effect (Fig. 6). Overall, these results are consistent with a recent origin for the species in the class of "zoonotic" ancient microbes. The finding that ancestry components found in pastoralist populations from the Steppe (CHG, EHG), but not those of Neolithic farming populations (LVN), are positively associated with "zoonotic" incidence, suggests that the cultural practices and living conditions of the former might have been more conducive to the emergence of novel zoonotic pathogens.

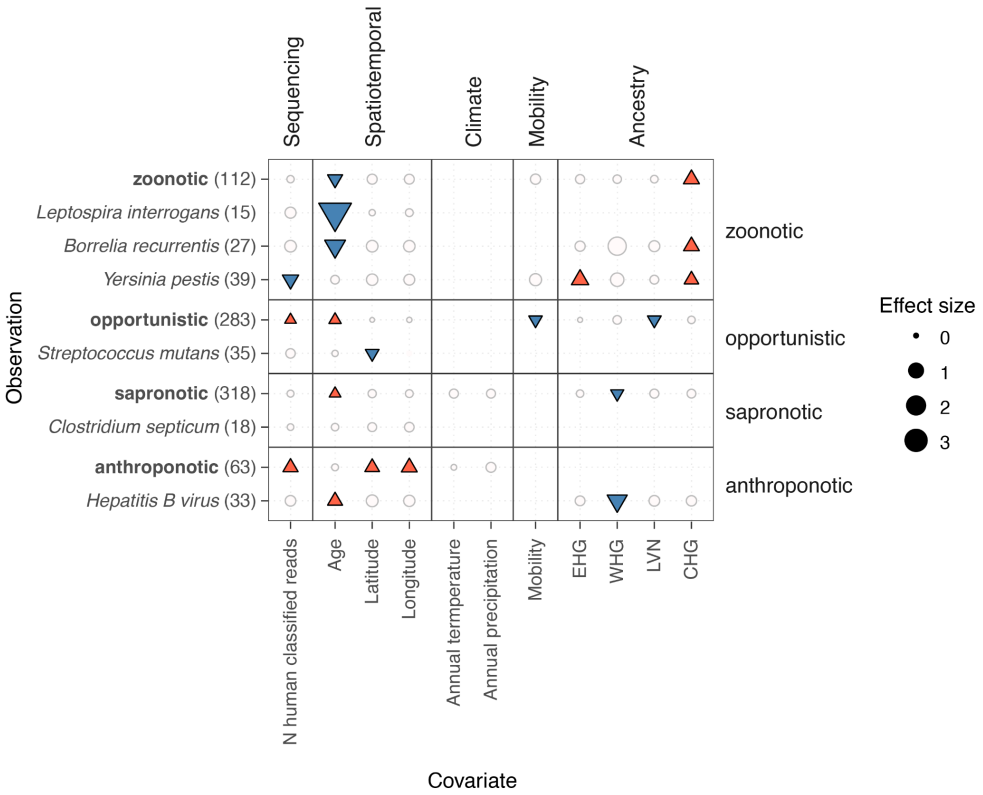



**Fig. 6. Predictors of ancient microbial species incidence.** Matrix showing effect sizes of 11 potential explanatory factors (columns) for presence of selected individual ancient microbial species (italics) or combined species classes (bold) inferred from spatiotemporal modelling. The total number of hits detected per species and class are given in brackets. For each species and class, the model with lowest Watanabe–Akaike information criterion is shown. Symbols indicate the covariates included in the respective model. Covariates with positive effect (2.5% and 97.5% posterior quantiles both positive) are shown as red triangles, whereas covariates with negative effect (2.5% and 97.5% posterior quantiles both negative) are shown as blue inverted triangles. Covariates included in the best-fitting model but without effect (posterior quantile range spanning zero) are indicated using white circles. Posterior mean of effect sizes is indicated by symbol size.

## Conclusions

The Holocene saw significant changes in global lifestyles, with agriculture, husbandry, and pastoralism becoming key practices. These changes have long been hypothesised to have favoured the emergence and spread of infectious diseases in a so-called "first epidemiological transition"[2]. In this study, we directly tested this hypothesis by carrying out the first large-scale characterization of the ancient pathogen landscape across Eurasia during the last ~12,500 years. Our results provide direct evidence of an epidemiological transition, with a first occurrence of zoonotic pathogens around 6,500 years ago, and consistent detection after 6,000 years ago. As our dataset is lacking information on RNA viruses, our zoonotic disease analyses underestimate the overall zoonotic disease burden; however, the timing is likely accurate as the conditions favouring zoonotic transmission of RNA viruses are similar to those of other pathogens. Although zoonotic cases also existed before the transition, the risk and extent of zoonotic pathogen transmission likely increased with the adoption of more widespread husbandry practices and pastoralism. Today, zoonoses are estimated to account for over 60% of newly emerging infectious diseases[56].

Strikingly, some of the highest detection rates were seen around ~5,000 BP, a period which saw major demographic turnover in Europe due to the migrations of Steppe pastoralist populations[3,4]. The observed association between zoonotic disease incidence and the CHG/EHG ancestry characteristic of those populations, but not with the LVN ancestry characteristic of the first agriculturists, indicates that the Neolithic to Bronze Age transition possibly played a pivotal role in the emergence and spread of zoonotic diseases in humans. It is possible that the Steppe pastoralists, through their long-term continuous exposure to animals, might have developed some immunity to certain zoonoses and that their dispersals carried these diseases westward and eastwards. Consequently, the genetic upheaval in Europe could have been facilitated by epidemic waves of zoonotic diseases sweeping through the continent. This may have caused significant population declines with depopulated areas repopulated by opportunistic settlers who intermixed with the remaining original population. Such a scenario would mirror the population decline of Indigenous people in the Americas following their exposure to Eurasian diseases introduced by European colonists[57,58]. Our findings also support the interpretation of increased pathogen pressure as a driver of positive selection on immunogenetic variants associated with risk of the autoimmune disease multiple sclerosis in Steppe populations around 5,000 years ago[59], and genetic adaptations in genes involved in host-pathogen interactions having occurred predominantly after the onset of the Bronze Age in Europe[60].

Importantly, expanding our analyses to the wider pathogen landscape allowed us to infer and contrast patterns of incidence between different species and types of pathogens. If ancient pathogen DNA of a species of interest is not detected in a particular region or time period, asserting whether this is due to the low disease incidence or confounding factors such as differential DNA preservation between different periods and environments is challenging when only distributions of single species are analysed. Our analyses counter these limitations; we demonstrate that pathogens with known epidemic potential and high detection rates, such as *Yersinia pestis* and *Borrelia recurrentis,* show striking differences in their detection rate over time, suggesting that low detection rate in these cases represent an actual reduction in the underlying incidence of the disease. During the early period, the incidence for plague matched those expected for an endemic disease, with continuous detection between ~5,700-2,700 years ago. The pattern in the later periods is consistent with an epidemic period, with distinct waves matching the historically described plague pandemics sweeping through Europe separated by periods without detection of the pathogen. This shift from endemic to epidemic is concurrent with significant changes in the *Y. pesti*s genome, particularly with regards to increased flea-transmissibility and pathogenicity[15,17]. The pattern for *B. recurrentis*, the causative agent of louse-borne relapsing fever, is almost entirely the opposite, with narrow peaks and long periods without detection suggesting local epidemics before ~2,700 years ago, and more widespread and consistent detection in recent periods. This later endemicity of louse-borne relapsing fever could be driven by human and environmental factors known to increase the risk of louse infestation, e.g., crowding and poor hygiene, further exacerbated by human disasters created by war, forced migrations, poverty, and famine[61].

Overall, our findings represent a first example of how to use the nascent field of genomic paleoepidemiology to decipher a map of the spatial and temporal distribution of diverse human bacteria, virus, and parasites over millennia. This map will develop as more ancient specimens are investigated, as will our abilities for matching their distribution with human genome data and archaeological and environmental data across human history. Our current map provides robust evidence that the Holocene

lifestyle changes resulted in an epidemiological transition to an increased zoonotic infectious disease burden, which has profoundly impacted global human health and history throughout the millennia and continues today.

## Methods

### Dataset

We compiled a dataset of ancient DNA shotgun-sequencing data from 1,313 ancient individuals, of which 130 are newly reported in this study. To facilitate ancient microbial DNA authentication, we excluded sequencing libraries subjected to UDG treatment which removes characteristic aDNA damage patterns from further analyses. Samples sequenced across multiple libraries were combined into single analysis units to maximise sensitivity for detection of ancient microbial DNA present in low abundance.

### Ancient microbial DNA screening

We carried out screening for ancient microbial DNA using a computational workflow combining kmer-based taxonomic classification, read mapping and ancient DNA authentication. We performed taxonomic classification of the sequencing reads using *KrakenUniq*[36], against a comprehensive database of complete bacterial, archaeal, viral, protozoan genomes in the RefSeq database. For each genus of interest showing ≥ 50 unique *k*-mers assigned, all sequencing reads classified were collected and aligned in parallel against a representative reference assembly for each individual species within the genus. We selected the assembly with the most unique *k*-mers assigned as the representative reference genome for each species in a particular sample. If no reads were assigned to any assembly of the species in *KrakenUniq*, we selected the first assembly for mapping. Read mapping was carried out using *minimap2*[38], using the '-sr' preset for mapping short read, and using *k*-mer size k=15 and minimizer window size w=5 for indexing the reference assemblies to increase sensitivity for potentially short and damaged aDNA sequencing reads. Mapped BAM files were subjected to duplicate removal using *picard MarkDuplicates*, and filtered for mapping quality MQ≥20. Ancient DNA damage rates were estimated using *metaDMG*[62].

### Authentication of ancient microbial DNA

To authenticate ancient microbial DNA, we calculated sets of summary statistics quantifying expected molecular characteristics of true positive ancient microbial DNA hits[63]:

#### *Similarity to the reference assembly*

Summary statistics in this category measure how similar sequencing reads are to a particular reference assembly, with true positive hits expected to show higher similarity than false positive hits. Summary statistics used include:

<u>Average edit distance</u>
The average number of mismatches in sequencing reads mapped to a particular reference (lower - more similar to reference)

<u>Average nuclear identity (ANI)</u>
The average number of bases in a mapped sequencing read matching the reference assembly, normalised by the read length (higher - more similar to reference).

Average number of soft-clipped bases
The average number of bases masked with soft-clipping (i.e., not aligned to reference) in a mapped sequencing read (lower - more similar to reference)

Number of unique *k*-mers assigned
The number of unique *k*-mers assigned to a particular reference assembly from *KrakenUniq* classification (higher - more similar to reference).

*Ancient DNA characteristics*

Summary statistics in this category measure the evidence for sequencing reads deriving from an ancient DNA source. Summary statistics used include:

Average read length
The average length in base pairs of sequencing reads mapped to a particular reference (shorter - more likely ancient)

Terminal aDNA substitution rates
The frequency of C>T (G>A) substitutions observed at the 5' (3') terminal base across all sequencing reads mapped to a particular reference (higher - more likely ancient)

Bayesian $D_{max}$
Bayesian estimator of ancient DNA damage rate from *metaDMG* ((higher - more likely ancient)

Bayesian $Z_x$
Bayesian estimator of significance of evidence for ancient DNA damage rate from *metaDMG* ((higher - more likely ancient)

*Evenness of genomic coverage*

Summary statistics in this category measure how evenly mapped sequencing reads are distributed across a reference assembly. Summary statistics used include:

Average read depth
The average number of reads covering a base in the reference assembly.

Breadth of coverage
The fraction of the reference assembly that is covered by one or more sequencing reads.

Expected breadth of coverage
Breadth of coverage expected for a particular average read depth, calculated[64] as

$$1 - e^{-(\text{average read depth})}$$

Ratio of observed over expected breadth of coverage

The ratio of breadth of coverage observed in mapping over breadth of coverage expected given observed average read depth (higher - more even coverage)

Relative entropy of read start positions

A measure for the information content of the genomic positions of mapped reads. To obtain this statistic, we calculate the frequency of read alignments with their start positions falling within 100bp windows along the reference assembly. The obtained frequency vector is converted into Shannon's entropy, and normalised using the maximum entropy attainable if the same total number of reads were evenly distributed across the 100bp windows (higher - more even coverage).

### Filtering of putative ancient microbial hits

From this initial screening, we then selected a subset of putative microbial "hits" (sample/species combinations) for further downstream analysis based on a set of aDNA authentication summary statistics:

- Number of mapped reads $\geq 20$
- 5' C>T deamination rate $\geq 0.02$
- Ratio of observed/expected breadth of coverage $\geq 0.3$
- Average number of soft-clipped bases in alignment $< 10$
- Rank of number of unique k-mers assigned $\leq 3$

For this initial filtered list of putative microbial hits, we ran *metaDMG* using the full Bayesian inference method to obtain Z-scores measuring the strength of evidence for observing aDNA damage.
The final list of putative individual ancient microbial hits was then obtained using the filtering cutoffs

- *metaDMG* Bayesian Dmax $\geq 0.03$
- *metaDMG* Bayesian Z $\geq 0.5$
- Ratio of observed/expected breadth of coverage $\geq 0.5$
- Relative entropy of read start positions $\geq 0.85$,
- Rank of number of unique k-mers assigned == 1

The result of this filtering is a single best-matching species hit for each sample and genus of interest. We note that this approach will miss potential cases where ancient DNA from multiple species of the same genus is present in the sample. However, due to the considerable challenges involved in distinguishing this scenario from false positives due to cross-mapping of ancient reads from a single source of DNA to reference assemblies of a closely related species (e.g. *Yersinia pestis / Yersinia pseudotuberculosis*), we opted for the conservative option of retaining only the best hit for each genus.

To increase sensitivity for ancient viral DNA, we re-ran the workflow on a viral-specific database of complete viral genomes and neighbour assemblies from RefSeq (https://www.ncbi.nlm.nih.gov/genome/viruses/about/assemblies/), using all reads classified as non-human from the previous run. We used the same filtering cutoffs described above, except for a lower cutoff for relative entropy of read start positions (0.6) for short viral genomes (< 10kb).

### Topic model analysis

We carried out topic model analysis on taxonomic classification profiles for each sample using the R package *fastTopics*[65] (https://github.com/stephenslab/fastTopics). We used the number of unique *k*-mers assigned to non-human genera from *KrakenUniq* as the observed count data for each sample, excluding genera with less than 50 unique *k*-mers assigned.

### Time series

To infer temporal dynamics of ancient microbial species, we calculated detection rates in a sliding window of k=21 temporally consecutive samples across the entire timeline of the 1,266 samples with dating information. For individual species, the detection rate for each window corresponds to the proportion of the 21 samples in each window that were positive for the species of interest. For analyses of species combined in classes, we calculated the detection rate as the ratio of the total number of hits within a class in the window over the total number of possible hits across all species in a window (21 samples x 258 species across all classes). For individual species with $n \geq 20$ hits or combined species classes, we further performed Bayesian change-point detection and time series decomposition (BEAST) [52] implemented in the R package *Rbeast* (https://github.com/zhaokg/Rbeast), using the detection frequencies described above as response variables.

### Spatiotemporal models of species incidence

To identify possible drivers of the observed spatiotemporal ancient microbial incidence, we combined the individual microbial species and the combined species groups with palaeoclimatic variables, human mobility estimates and kriged estimates of ancestry composition for Holocene West Eurasia. Palaeoclimatic reconstructions were accessed using the CHELSA-Trace21k data, which provides global monthly climatologies for temperature and precipitation at 30 arcsec spatial resolution in 100-year time steps for the last 21,000 years[66]. To pair the microbial species/groups to the palaeoclimatic reconstructions, we took the average climatic value across all the time steps that fall within the microbial species/group age ± sd at each of the sampling locations. Palaeoclimatic variables considered were annual mean temperature (BIO01) and annual precipitation (BIO12). Human mobility values were accessed from Schmid & Schiffels[67] and approximately represent the distance in kilometres (km) between the burial location of the ancient human individual and its putative ancestral origin, based on patterns of genetic similarity derived from a MDS analysis. Microbial species/groups were paired to the mobility estimate of the ancient human individual that occurs closest in space and time. Kriged ancestry estimates were extracted from Allentoft *et al.*[68], using the spatiotemporal ancestry kriging method from Racimo et al.[69], and paired to the closest spatiotemporal location of the ancient human remain where the corresponding microbial species/groups were sampled.

To determine the influence of the covariates on the microbial incidence, we used a hierarchical Bayesian model implemented in the *inlabru* R package[55,70], where ancient microbial presence/absence follows a binomial distribution and the spatiotemporal variables (latitude, longitude and sample age), number of human-classified reads, palaeoclimatic variables, human mobility and human ancestry constitute the linear predictors. All covariates were normalized before the analyses. For each microbial species and group, we tested multiple models with different sets of covariates: 1) palaeoclimate + mobility + ancestry, 2) palaeoclimate + mobility, 3) palaeoclimate + ancestry, 4) only climate, 5) mobility + ancestry, 6) only

mobility, 7) only ancestry, 8) no climate, nor mobility, nor ancestry. Spatiotemporal variables and number of human-classified reads were included in all models. We used the Watanabe-Aikaike Information Criterion (WAIC) to assess the model fit to each set of covariates. The results shown are only for the best-performing models (i.e., models with the lowest WAIC score for each microbial species or combined species group).

## Data availability

All raw sequence data used in this study is available as trimmed read files (FASTQ) at the European Nucleotide Archive under accession PRJEB65256. Processed analysis files including *KrakenUniq* database file and metagenomic profiling results, microbial species read alignments (BAM format) as well as per-sample summary tables and plots from screening pipeline are available at Zenodo under accession XX.

## Code availability

A *snakemake* workflow implementing the computational screening pipeline is available at https://github.com/martinsikora/pathopipe.

## Acknowledgements

The Lundbeck Foundation GeoGenetics Centre is supported by the the Lundbeck Foundation (R302-2018-2155, R155-2013-16338), the Novo Nordisk Foundation (NNF18SA0035006), the Wellcome Trust (UNS69906), Carlsberg Foundation (CF18-0024), the Danish National Research Foundation (44113220) and the University of Copenhagen (KU2016 programme). Additional support was provided by Germany's Excellence Strategy (EXC- 2077), project 390741603 "The Ocean Floor – Earth's Uncharted Interface". We thankfully acknowledge Illumina Inc. for collaboration. EW thanks St. John's College, Cambridge, for providing a stimulating environment of discussion and learning. This work was further supported by the Swedish Foundation for Humanities and Social Sciences grant (Riksbankens Jubileumsfond M16-0455:1) to KK. M.E.A. was supported by Marie Skłodowska-Curie Actions of the EU (grant no. 300554), The Villum Foundation (grant no. 10120) and Independent Research Fund Denmark (grant no. 7027-00147B). G.S. is funded by Marie Skłodowska-Curie Individual Fellowship 'PALAEO-ENEO' (grant agreement number 751349). F. R. is supported by a Villum Young Investigator Grant (project no. 00025300), a Novo Nordisk Fonden Data Science Ascending Investigator Award (NNF22OC0076816) and by the European Research Council (ERC) under the European Union's Horizon Europe programme (grant agreements No. 101077592 and 951385). N.O., R.Å., L.H. and B.N. are financially supported by Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. A. I. and L. F. thank the OAK Foundation.

## References

1. Lewis, C. M., Jr, Akinyi, M. Y., DeWitte, S. N. & Stone, A. C. Ancient pathogens provide a window into health and well-being. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2209476119 (2023).

2. Barrett, R., Kuzawa, C. W., McDade, T. & Armelagos, G. J. EMERGING AND RE-EMERGING INFECTIOUS DISEASES: The third epidemiologic transition. *Annu. Rev. Anthropol.* **27**, 247–271 (1998).
3. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
4. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
5. Roser, M., Ritchie, H. & Dadonaite, B. Child and Infant Mortality. *Our World in Data* (2013).
6. Dobyns, H. F. Disease transfer at contact. *Annu. Rev. Anthropol.* **22**, 273–291 (1993).
7. *Plague and the end of antiquity: The pandemic of 541-750*. (Cambridge University Press, 2008).
8. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* **15**, 379–393 (2014).
9. Quintana-Murci, L. Human Immunology through the Lens of Evolutionary Genetics. *Cell* **177**, 184–199 (2019).
10. Page, A. E. *et al.* Reproductive trade-offs in extant hunter-gatherers suggest adaptive mechanism for the Neolithic expansion. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4694–4699 (2016).
11. Rascovan, N. *et al.* Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. *Cell* **176**, 295–305.e10 (2019).
12. Fuchs, K. *et al.* Infectious diseases and Neolithic transformations: Evaluating biological and archaeological proxies in the German loess zone between 5500 and 2500 BCE. *Holocene* **29**, 1545–1557 (2019).
13. Abegg, C., Desideri, J., Dutour, O. & Besse, M. More than the sum of their parts: reconstituting the paleopathological profile of the individual and commingled Neolithic populations of Western Switzerland. *Archaeol. Anthropol. Sci.* **13**, (2021).
14. Bos, K. I. *et al.* A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506–510 (2011).
15. Rasmussen, S. *et al.* Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. *Cell* **163**,

571–582 (2015).

16. Andrades Valtueña, A. *et al.* The Stone Age Plague and Its Persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691.e8 (2017).
17. Spyrou, M. A. *et al.* Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234 (2018).
18. Keller, M. *et al.* Ancient genomes from across Western Europe reveal early diversification during the First Pandemic (541-750). *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12363–12372 (2019).
19. Susat, J. *et al.* A 5,000-year-old hunter-gatherer already plagued by Yersinia pestis. *Cell Rep.* **35**, 109278 (2021).
20. Feldman, M. *et al.* A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).
21. Andrades Valtueña, A. *et al.* Stone Age genomes shed light on the early evolution, diversity, and ecology of plague. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2116722119 (2022).
22. Guellil, M. *et al.* A genomic and historical synthesis of plague in 18th century Eurasia. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28328–28335 (2020).
23. Bos, K. I. *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
24. Vågene, Å. J. *et al.* Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. *Nat. Commun.* **13**, 1195 (2022).
25. Duggan, A. T. *et al.* 17th Century Variola Virus Reveals the Recent History of Smallpox. *Curr. Biol.* **26**, 3407–3412 (2016).
26. Mühlemann, B. *et al.* Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* **369**, (2020).
27. Mühlemann, B. *et al.* Ancient Hepatitis B viruses from the Bronze Age to the Medieval.
28. Kocher, A. *et al.* Ten millennia of hepatitis B virus evolution. *Science* **374**, 182–188 (2021).
29. Krause-Kyora, B. *et al.* Neolithic and medieval virus genomes reveal complex evolution of hepatitis

B. *Elife* **7**, (2018).

30. Guellil, M. *et al.* Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10422–10427 (2018).
31. Schuenemann, V. J. *et al.* Genome-Wide Comparison of Medieval and Modern Mycobacterium leprae. *Science* **341**, 179–183 (2013).
32. Guellil, M. *et al.* Ancient herpes simplex 1 genomes reveal recent viral structure in Eurasia. *Sci Adv* **8**, eabo4435 (2022).
33. van Dorp, L. *et al.* Plasmodium vivax Malaria Viewed through the Lens of an Eradicated European Strain. *Mol. Biol. Evol.* **37**, 773–785 (2020).
34. Key, F. M. *et al.* Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. *Nat Ecol Evol* **4**, 324–333 (2020).
35. Mühlemann, B. *et al.* Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7557–7562 (2018).
36. Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 198 (2018).
37. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45 (2016).
38. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
39. Hsu, T. *et al.* Comparative genomics and genome biology of. *Emerg. Microbes Infect.* **8**, 827–840 (2019).
40. Spyrou, M. A. *et al.* Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. *Nat. Commun.* **10**, 4470 (2019).
41. Neumann, G. U. *et al.* Ancient Yersinia pestis and Salmonella enterica genomes from Bronze Age Crete. *Curr. Biol.* **32**, 3641–3649.e8 (2022).
42. Swali, P. *et al.* Yersinia pestis genomes reveal plague in Britain 4000 years ago. *Nat. Commun.* **14**,

2930 (2023).

43. Warrell, D. A. Louse-borne relapsing fever (Borrelia recurrentis infection). *Epidemiol. Infect.* **147**, e106 (2019).

44. Pfrengle, S. *et al.* Mycobacterium leprae diversity and population dynamics in medieval Europe from novel ancient genomes. *BMC Biol.* **19**, 220 (2021).

45. Steinegger, M. & Salzberg, S. L. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* **21**, 115 (2020).

46. Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M. & Snow, R. W. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect. Dis.* **4**, 327–336 (2004).

47. Hulden, L. & Hulden, L. The decline of malaria in Finland--the impact of the vector and social variables. *Malar. J.* **8**, 94 (2009).

48. de-Dios, T. *et al.* Genetic affinities of an eradicated European strain. *Microb Genom* **5**, (2019).

49. Singer, M., Bulled, N., Ostrach, B. & Mendenhall, E. Syndemics and the biosocial conception of health. *Lancet* **389**, 941–950 (2017).

50. Zasada, A. A., Zaleska, M., Podlasin, R. B. & Seferynska, I. The first case of septicemia due to nontoxigenic Corynebacterium diphtheriae in Poland: case report. *Ann. Clin. Microbiol. Antimicrob.* **4**, 8 (2005).

51. Han, X. Y., Tarrand, J. J., Dickey, B. F. & Esteva, F. J. Helicobacter pylori bacteremia with sepsis syndrome. *J. Clin. Microbiol.* **48**, 4661–4663 (2010).

52. Zhao, K. *et al.* Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sens. Environ.* **232**, 111181 (2019).

53. Demeure, C. *et al.* Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination and diagnostics. *Microbes Infect.* **21**, 202–212 (2019).

54. Hubálek, Z. Emerging human infectious diseases: anthroponoses, zoonoses, and sapronoses. *Emerg. Infect. Dis.* **9**, 403–404 (2003).

55. doi:10.1111/2041-210X.13168.

56. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).

57. Meltzer, D. J. *First Peoples in a New World: Colonizing Ice Age America*. (Univ of California Press, 2009).

58. Collen, E. J., Johar, A. S., Teixeira, J. C. & Llamas, B. The immunogenetic impact of European colonization in the Americas. *Front. Genet.* **13**, 918227 (2022).

59. Barrie, W. *et al.* Genetic risk for Multiple Sclerosis originated in Pastoralist Steppe populations. (2022) doi:10.1101/2022.09.23.509097.

60. Kerner, G. *et al.* Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell Genom.* **3**, 100248 (2023).

61. Cutler, S. J. Relapsing fever--a forgotten disease revealed. *J. Appl. Microbiol.* **108**, 1115–1122 (2010).

62. Michelsen, C. *et al.* MetaDMG – A fast and accurate ancient DNA damage toolkit for metagenomic data. *bioRxiv* (2022) doi:10.1101/2022.12.06.519264.

63. Warinner, C. *et al.* A Robust Framework for Microbial Archaeology. *Annu. Rev. Genomics Hum. Genet.* **18**, 321–356 (2017).

64. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).

65. Carbonetto, P., Sarkar, A., Wang, Z. & Stephens, M. Non-negative matrix factorization algorithms greatly improve topic model fits. (2021) doi:10.48550/ARXIV.2105.13440.

66. Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H. & Zimmermann, N. E. CHELSA-TraCE21k -- high-resolution (1 km) downscaled transient temperature and precipitation data since the Last Glacial Maximum. *Clim. Past* **19**, 439–456 (2023).

67. Schmid, C. & Schiffels, S. Estimating human mobility in Holocene Western Eurasia with large-scale ancient genomic data. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2218375120 (2023).

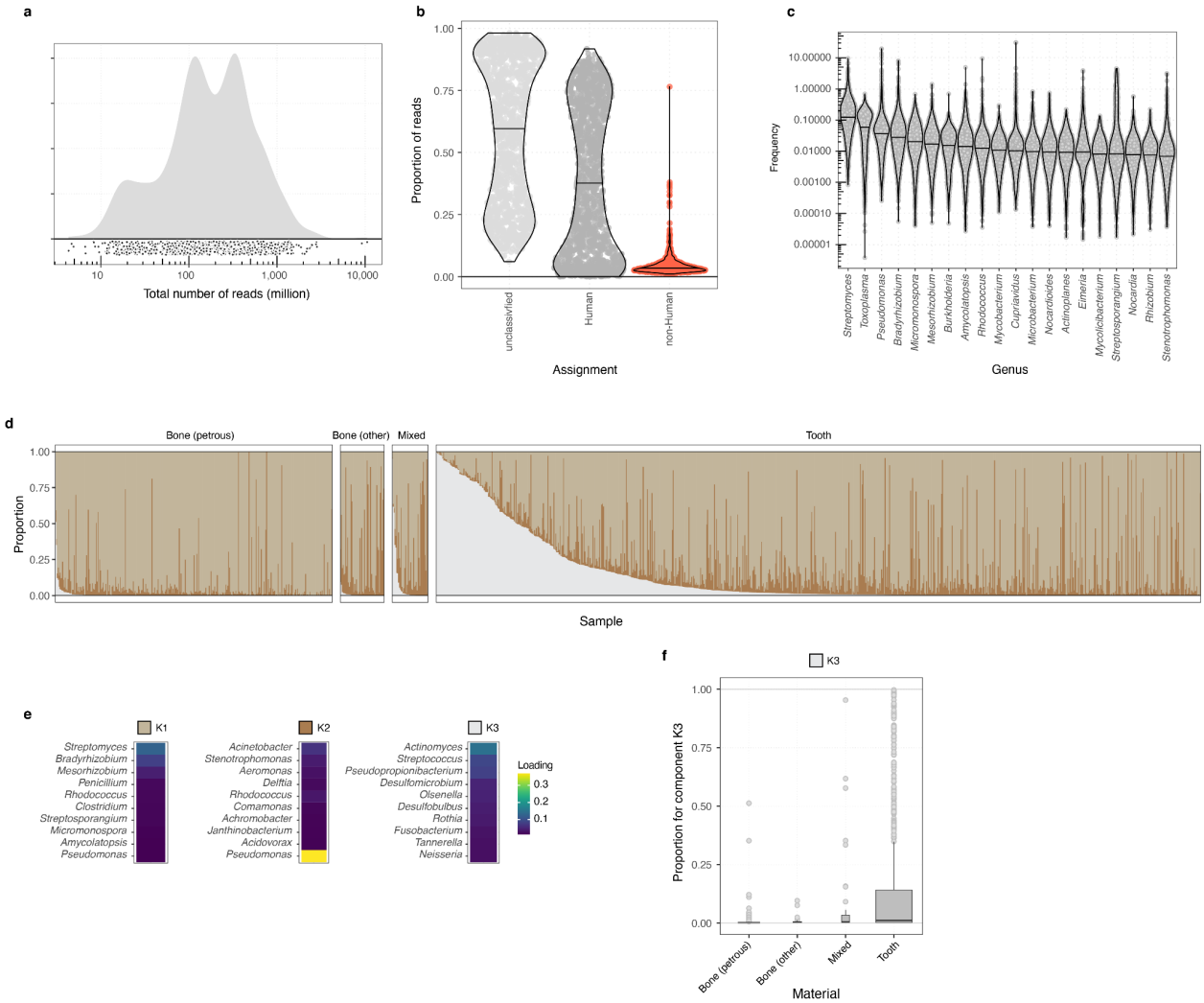
68. Allentoft, M. E. *et al.* Population genomics of Stone Age Eurasia. *bioRxiv* (2022)

doi:10.1101/2022.05.04.490594.

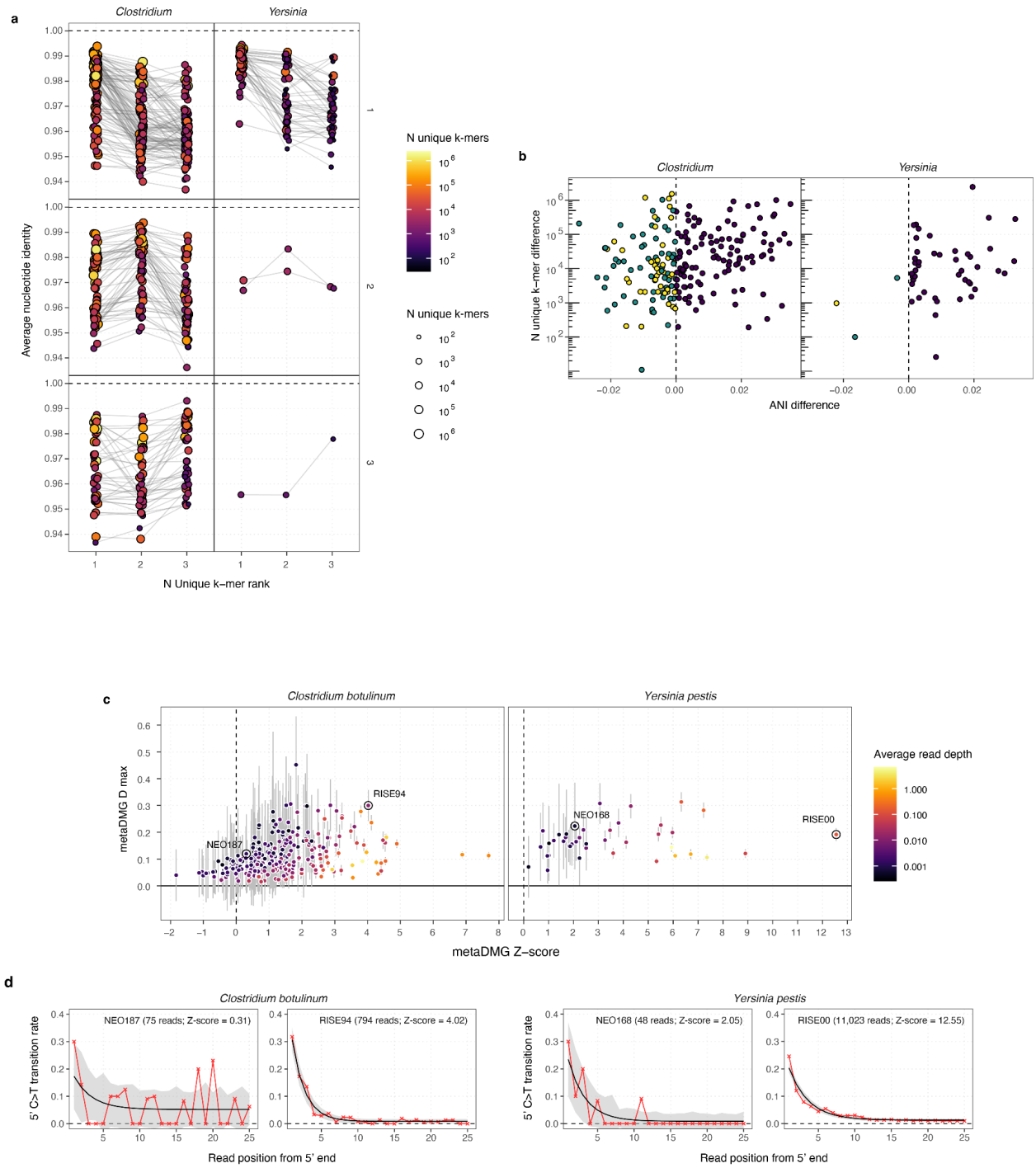
69. Racimo, F. *et al.* The spatiotemporal spread of human migrations during the European Holocene. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8989–9000 (2020).

70. Lindgren, F. & Rue, H. Bayesian Spatial Modelling withR-INLA. *J. Stat. Softw.* **63**, (2015).

## Extended Data Figures

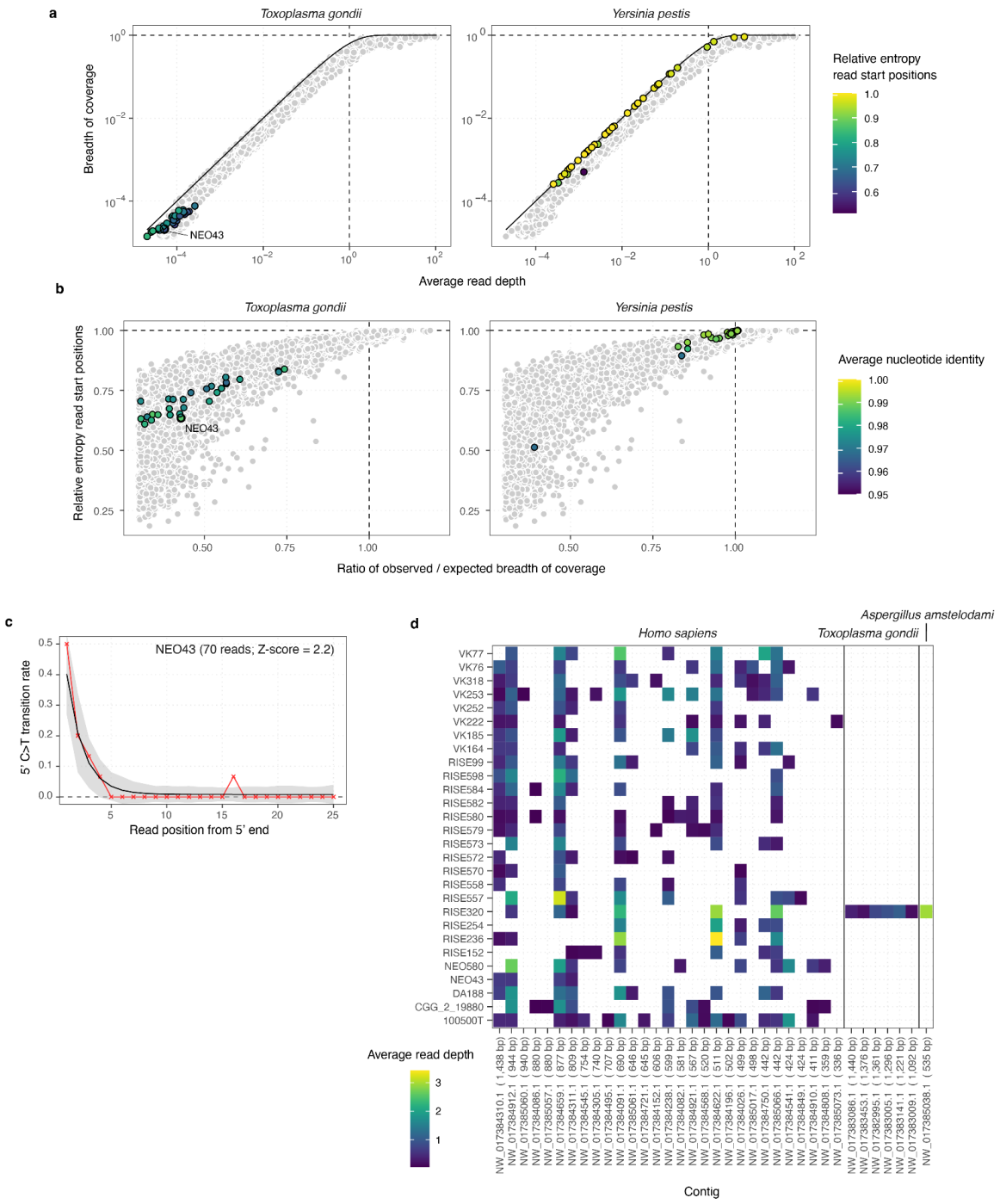



**Extended Data Fig. 1. Metagenome composition. a,** Distribution of total number of sequencing reads screened across the 1,313 study samples. **b,** Violin plots showing distributions of proportions of reads classified as human, non-human or not classified for the study samples. Median values for each genus are indicated by horizontal lines. **c,** Violin plots showing fraction of reads classified on the taxonomic level of genus, for the top 20 most abundant genera. **d,** Barplots showing inferred proportions for K=3 topics (indicated by fill colour) from topic model analysis for 1,272 study samples with sample material information. **e,** Factor loadings for the 10 highest loading genera for each of the K=3 topics from the topic model analysis. **f,** Boxplots showing distributions of proportions for topic K3 (associated with oral microbiome taxa) in different sample materials.




**Extended Data Fig. 2. Authentication of ancient microbial DNA.** Plots showing ancient microbial DNA authentication summary statistics characterising similarity of sequencing reads to the reference assembly (**a, b**), and ancient DNA damage patterns (**c, d**), for example hits from the genera *Clostridium* and *Yersinia.* **a,** Relationship of ANI of mapped reads and unique *k*-mer counts from *KrakenUniq*. Each sample with a hit in the indicated genus is represented by three points connected with a grey line, showing ANI for the species hit with the first, second and third highest number of unique *k*-mers observed. Panels in rows differentiate samples where the highest ANI of mapped reads was observed for species hit with the highest (top), second highest (middle) or third highest (bottom) unique *k*-mer count. **b,** Plot showing relationship of difference in unique *k*-mer count and ANI between the species hit with the highest unique *k*-mer count and the hit of the remaining two species showing highest ANI for the respective sample (coloured points). Fill colour indicates whether the highest ANI was observed in the species hit with highest (blue), second highest (green) or third highest (yellow) unique *k*-mer count. **c,** Plot showing Bayesian estimator of ancient DNA

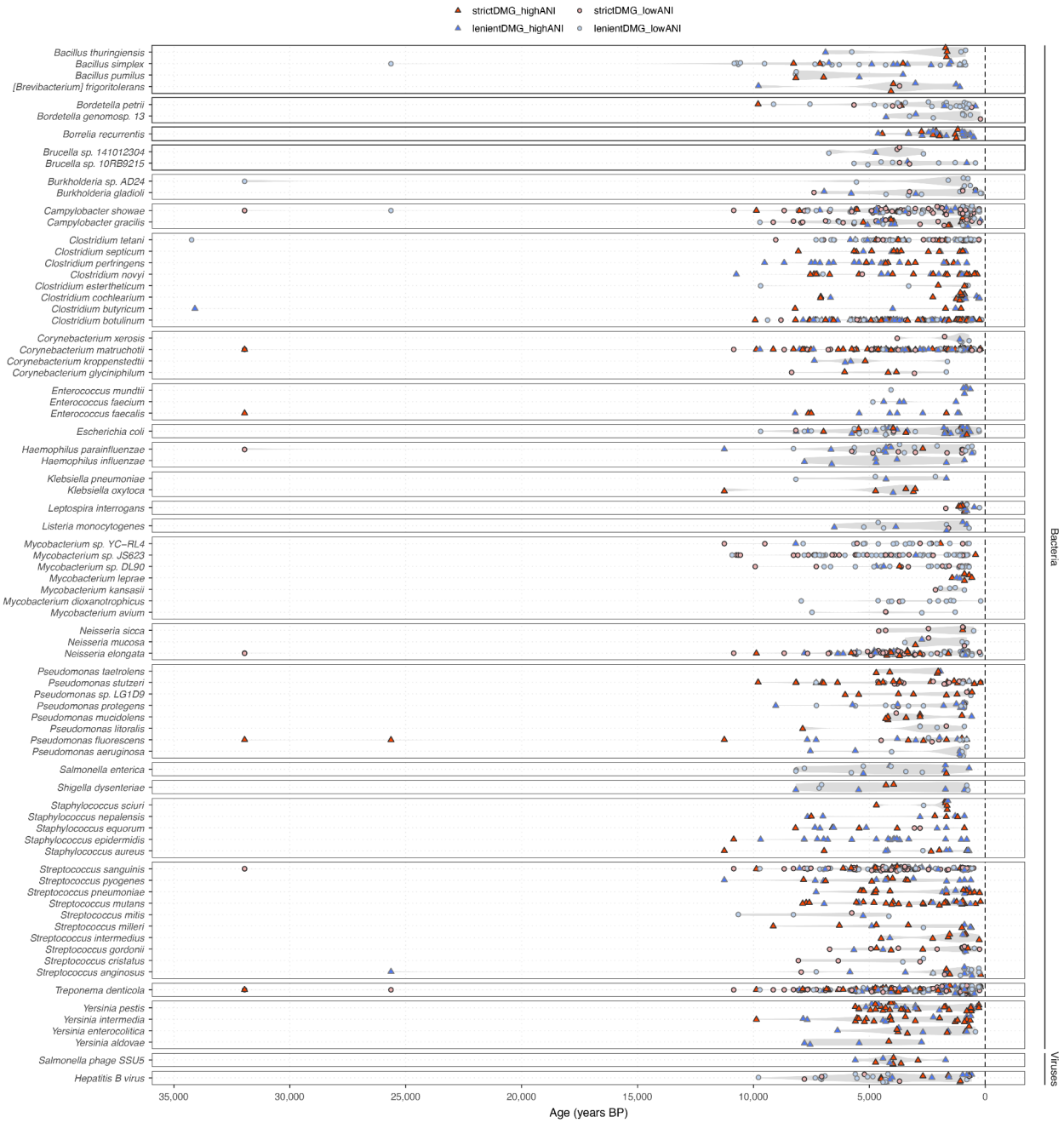
damage (D max) and significance (Z-score) obtained from *metaDMG,* for hits in species *Clostridium botulinum* (left) and *Yersinia pestis* (right). Error bars indicate ± 1 standard deviation, and symbol fill colour indicates average read depth for mapped reads. Samples used as examples in ancient DNA damage curves (**d**) are labeled and indicated with black circles. **d,** Ancient DNA damage patterns for four example hits in species *Clostridium botulinum* and *Yersinia pestis.* Plots show observed nucleotide misincorporation frequencies (red symbols and line) and *metaDMG* fit (black line) and 68% credible intervals (shaded region) for C>T transitions as a function of distance from the 5' read end.



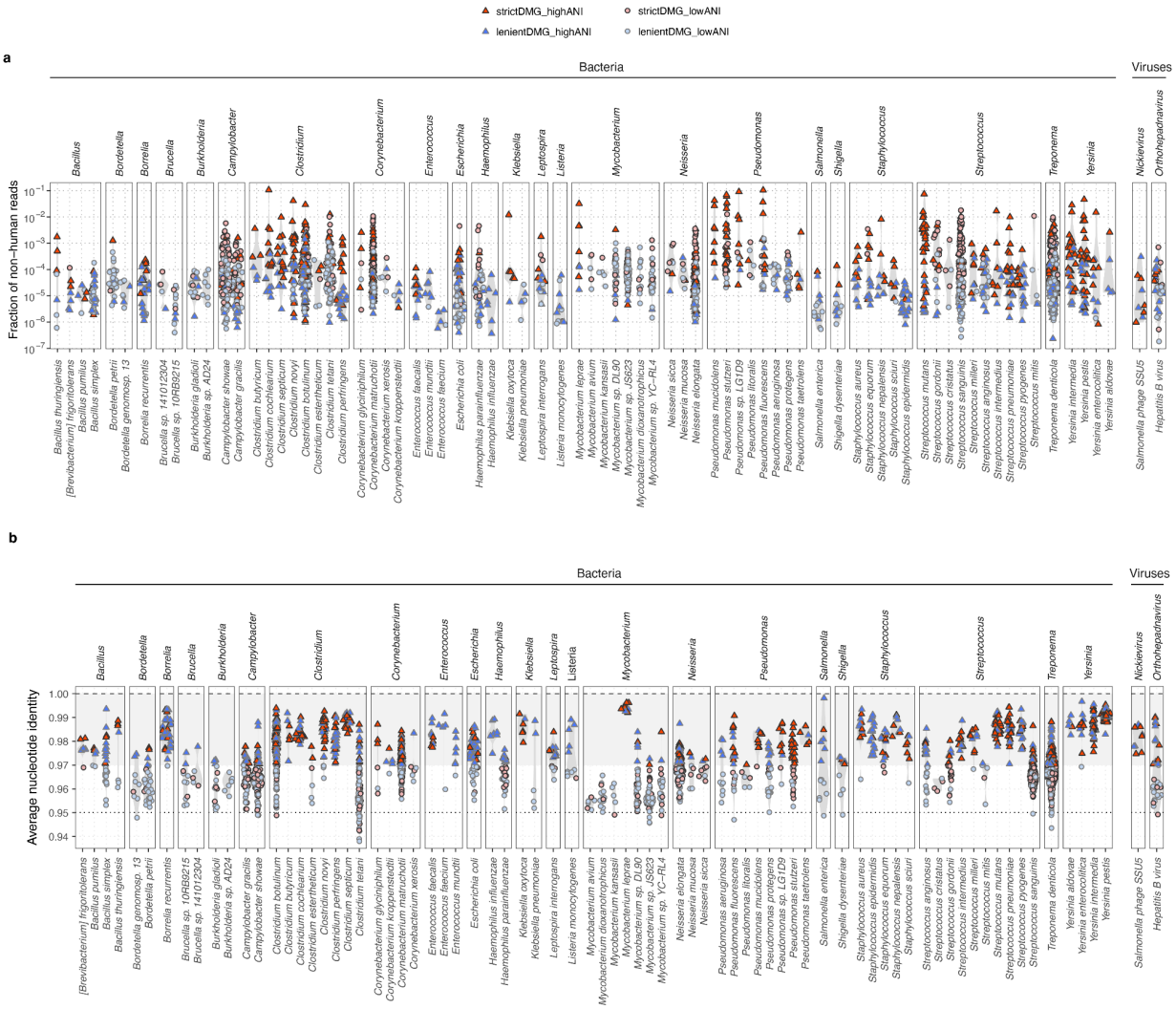
**Extended Data Fig. 3. Evenness of genomic coverage and reference assembly contamination. a,** Observed breadth of genomic coverage as a function of average read depth. Coloured symbols indicate hits in species *Toxoplasma gondii* (left panel) and *Yersinia pestis* (right panel), with symbol colour indicating relative entropy of read start positions (Methods). Grey symbols indicate background distribution of other species hits passing initial filters (Ratio of observed/expected breadth of coverage ≥ 0.3, Methods). Solid black line shows theoretical expected breadth of coverage for a given average read depth[64]. **b,** Relative entropy of read start positions as a function of the ratio of observed over expected breadth of coverage, for hits in species *Toxoplasma gondii* (left panel) and *Yersinia pestis* (right panel). Symbol colour indicates ANI for the respective sample. **c,** Ancient DNA damage patterns for example hit deriving from human DNA contamination, NEO43 (also indicated in **a, b**) in species *Toxoplasma gondii.* Shown are nucleotide misincorporation frequencies (red symbols and line) and *metaDMG* fit (black line) and 68% credible intervals (shaded region) for C>T transitions as a function of distance from the 5' read end. **d**, Heatmap showing average read depth in genomic contigs covered in samples with hits for species *Toxoplasma gondii* passing initial filters (ratio of observed/expected breadth of coverage ≥ 0.3, Methods). Accessions and length of contigs from reference assembly *Toxoplasma gondii* ME49 (GCF_000006565.2) are indicated, and contigs are grouped according to the species scoring highest in a BLAST search (megablast, nt database). All six contigs covered in sample RISE320 with *Toxoplasma gondii* as highest-scoring BLAST hit were matching sequences in the 28S ribosomal RNA.

Count
0
100
200
strictDMG highANI
lenientDMG highANI
strictDMG lowANI
lenientDMG lowANI
Bacillus simplex
Bacillus thuringiensis
[Brevibacterium] frigoritolerans
Bacillus pumilus
Bacillus butanolivorans
Bacillus clausii
Bacillus megaterium
Bacillus asahii
Bacillus mycoides
Bacillus toyoninis
Bacillus halodurans
Bacillus oceanisediminis
Bacillus sp. X1(2014)
Bacillus
Bordetella petrii
Bordetella genomosp. 13
Bordetella trematum
Bordetella bronchiseptica
Bordetella sp. HZ20
Bordetella
Borrelia recurrentis
Borrelia crocidurae
Borrelia
Brucella sp. 10RB9215
Brucella sp. 14I012304
Brucella sp. 09RB8471
Brucella vulpis
Brucella pinnipedialis
Brucella sp. 09RB8910
Brucella
Burkholderia gladioli
Burkholderia sp. AD24
Burkholderia sp. CLC4172
Burkholderia multivorans
Burkholderia
Campylobacter showae
Campylobacter gracilis
Campylobacter
Clostridium botulinum
Clostridium tetani
Clostridium novyi
Clostridium perfringens
Clostridium cochlearium
Clostridium septicum
Clostridium butyricum
Clostridium estertheticum
Clostridium chauvoei
Clostridium sp. BNL1100
Clostridium
Corynebacterium matruchotii
Corynebacterium glycinophilum
Corynebacterium xerosis
Corynebacterium kroppenstedtii
Corynebacterium jeikeium
Corynebacterium variabile
Corynebacterium striatum
Corynebacterium ammoniagenes
Corynebacterium glutamicum
Corynebacterium ursicoloriverans
Corynebacterium crudilactis
Corynebacterium diphtheriae
Corynebacterium maris
Corynebacterium segmentisum
Corynebacterium sphenisci
Corynebacterium urealyticum
Corynebacterium
Enterococcus faecalis
Enterococcus mundtii
Enterococcus faecium
Enterococcus casseliflavus
Enterococcus avium
Enterococcus cecorum
Enterococcus durans
Enterococcus hirae
Enterococcus sp. FDAARGOS_375
Enterococcus sp. CR-Ec1
Enterococcus
Escherichia coli
Escherichia
Haemophilus parainfluenzae
Haemophilus influenzae
Haemophilus
Helicobacter cinaedi
Helicobacter pylori
Helicobacter
Klebsiella oxytoca
Klebsiella pneumoniae
Klebsiella aerogenes
Klebsiella michiganensis
Klebsiella
Legionella pneumophila
Legionella anisa
Legionella longbeachae
Legionella
Leptospira interrogans
Leptospira borgpetersenii
Leptospira biflexa
Leptospira
Listeria monocytogenes
Listeria
Mycobacterium sp. JS623
Mycobacterium sp. DL90
Mycobacterium sp. YC-RL4
Mycobacterium ulcerans/marinum
Mycobacterium leprae
Mycobacterium avium
Mycobacterium kansasii
Mycobacterium paragordonae
Mycobacterium sp. EPA/15
Mycobacterium marseillense
Mycobacterium
Neisseria elongata
Neisseria sicca
Neisseria mucosa
Neisseria subflava
Neisseria meningitidis
Neisseria weaveri
Neisseria
Pseudomonas stutzeri
Pseudomonas protegens
Pseudomonas fluorescens
Pseudomonas aeruginosa
Pseudomonas sp. LG1D2
Pseudomonas mucidolens
Pseudomonas litoralis
Pseudomonas taetrolens
Pseudomonas brennneri
Pseudomonas putida
Pseudomonas sp. M30-35
Pseudomonas moraviensis
Pseudomonas sp. 02C 26
Pseudomonas sp. RU47
Pseudomonas symbiotica
Pseudomonas silesiensis
Pseudomonas sp. ACM7
Pseudomonas sp. CC8-YY-74
Pseudomonas sp. UMK017WK12111
Pseudomonas alcaligenes
Pseudomonas arsenicoxydans
Pseudomonas azotoformans
Pseudomonas psychrophila
Pseudomonas rhizosphaerae
Pseudomonas sp. ATCC 13867
Pseudomonas sp. FDAARGOS 380
Pseudomonas sp. S09Q 359
Pseudomonas
Rickettsia felis
Rickettsia
Salmonella enterica
Salmonella
Shigella dysenteriae
Shigella boydii
Shigella flexneri
Shigella
Staphylococcus epidermidis
Staphylococcus aquorum
Staphylococcus aureus
Staphylococcus nepalensis
Staphylococcus sciuri
Staphylococcus cohnii
Staphylococcus hominis
Staphylococcus saprophyticus
Staphylococcus succinus
Staphylococcus simulans
Staphylococcus
Streptococcus sanguinis
Streptococcus mutans
Streptococcus anginosus
Streptococcus pneumoniae
Streptococcus gordonii
Streptococcus pyogenes
Streptococcus intermedius
Streptococcus mitis
Streptococcus cristatus
Streptococcus mitis
Streptococcus agalactiae
Streptococcus dysgalactiae
Streptococcus oralis
Streptococcus gallolyticus
Streptococcus lutetiensis
Streptococcus parauberis
Streptococcus salivarius
Streptococcus sp. HSISM1
Streptococcus suis
Streptococcus thermophilus
Streptococcus
Treponema denticola
Treponema pallidum
Treponema
Vibrio anguillarum
Vibrio furnissii
Vibrio parahaemolyticus
Vibrio splendidus
Vibrio
Yersinia pestis
Yersinia intermedia
Yersinia enterocolitica
Yersinia aldovae
Yersinia aleksiciae
Yersinia ruckeri
Yersinia rohdei
Yersinia
Plasmodium vivax
Plasmodium
Primate erythroparvovirus 1
Erythroparvovirus
Escherichia virus DE3
Lambdavirus
Salmonella phage SSU5
Nickievirus
Hepatitis B virus
Orthohepadnavirus
Bacteria
Eukaryota
Viruses

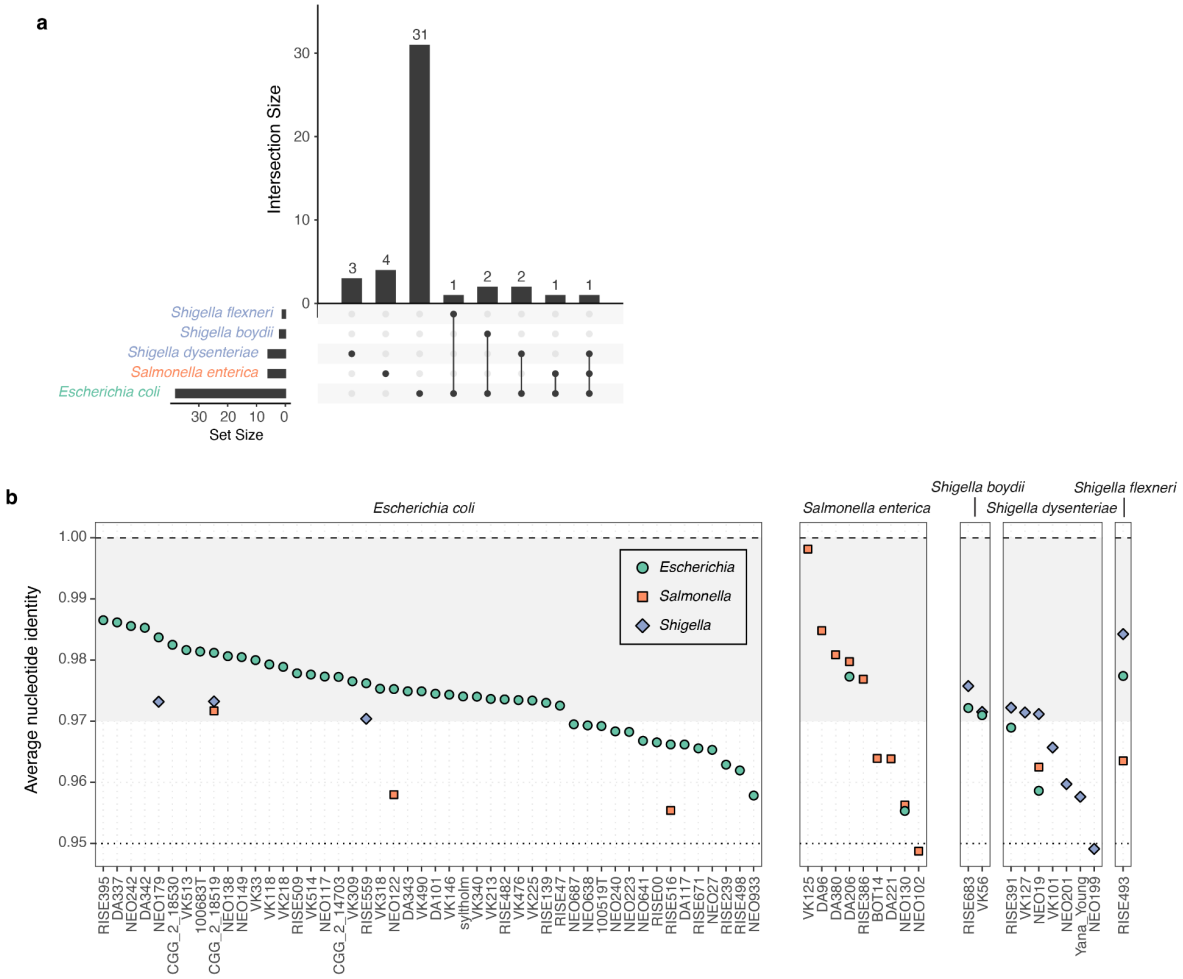
**Extended Data Fig. 4. Ancient microbial species detection rate.** Barplot showing total number of putative ancient microbial species hits (overall detection rate in brackets) for bacterial, eukaryotic and viral ($n \geq 4$) genera. Bar colour and shading distinguishes counts in the different ANI and aDNA damage categories.



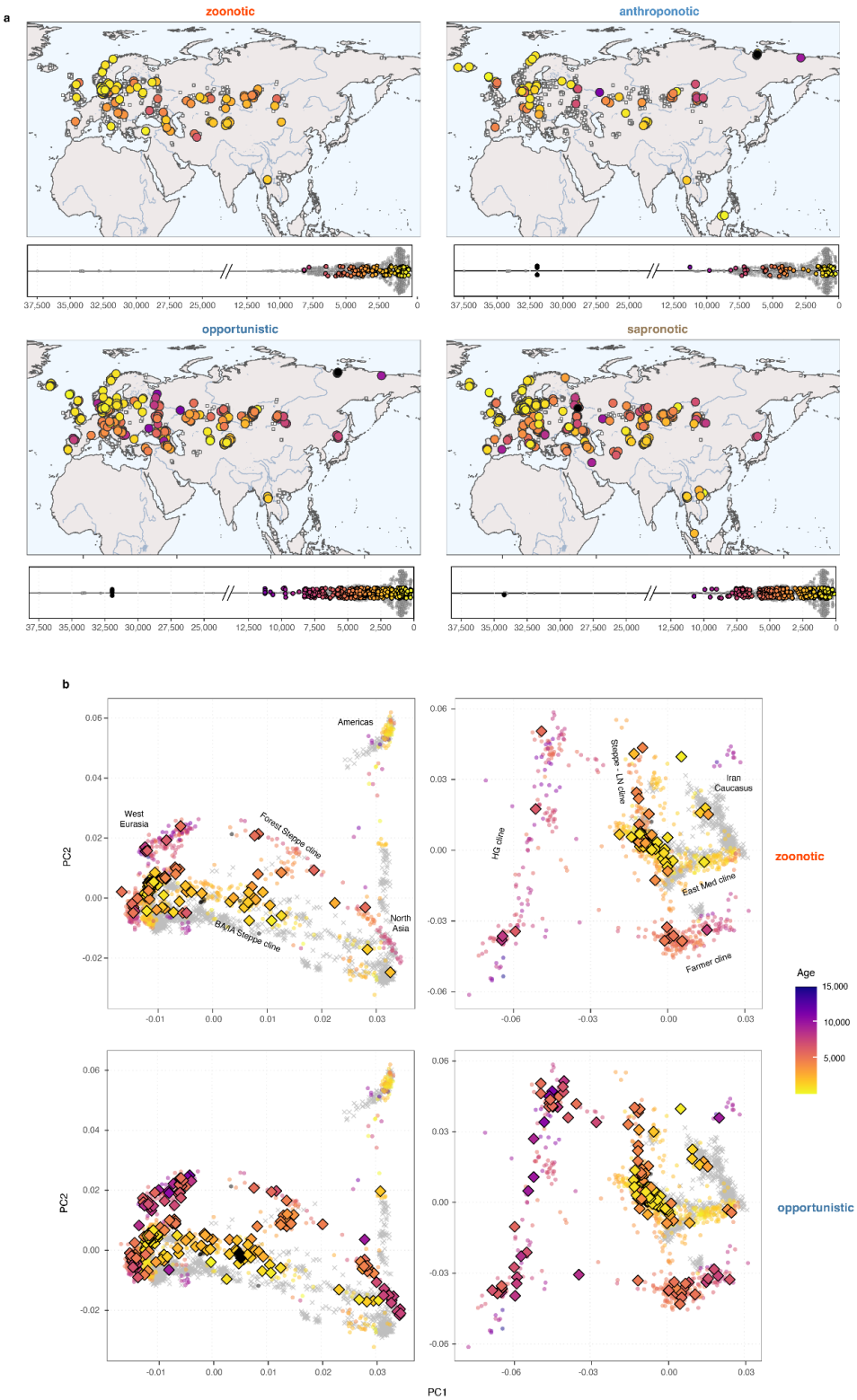

**Extended Data Fig. 5. Ancient microbial species timeline.** Distribution of sample ages for putative ancient microbial species detected in n ≥ 5 samples. Symbol shape and colour indicate species hit category.




**Extended Data Fig. 6. Ancient microbial DNA characteristics. a,** Distributions of relative abundance of ancient microbial DNA, measured as fraction of final mapped reads of total reads sequenced, for individual species hits detected in n ≥ 5 samples. **b**, Distributions of ANI for individual species hits detected in n ≥ 5 samples. Symbol shape and colour indicate species hit category.




**Extended Data Fig. 7. Ancient microbial hits within the closely related genera of *Escherichia*, *Shigella* and *Salmonella*. a,** UpSet plot showing the within-sample overlap of ancient microbial hits **b,** Distribution of ANI for ancient microbial hits. Samples are grouped according to species assignment for the ancient microbial hit with highest ANI, and sorted by ANI.




**Extended Data Fig. 8. Spatiotemporal distribution and host genetic structure for ancient microbial classes. a,** Panels showing geographic distributions (top) and timelines (bottom) for identified cases of ancient microbial hits in the zoonotic, anthroponotic, opportunistic and sapronotic classes (indicated by coloured circle). Geographic locations and age distributions of all 1,313 study samples are shown in each panel using white squares. **b,** Principal component analyses showing ancient and

modern human genetic population structure in non-African (left panels) and west Eurasian (right panels) individuals. grey crosses indicate present-day individuals, whereas coloured symbols indicate ancient individuals (coloured by sample age). Diamonds with black outlines indicate position in PCA space for samples with hits in zoonotic (top panels) or opportunistic (bottom panels) class. Major poles and clines of known ancient and modern human ancestry groups are indicated in the panels for zoonotic species hits.