# REVIEWS

# Ancient pathogen genomics as an emerging tool for infectious disease research

*Maria A. Spyrou* ✪ *, Kirsten I. Bos, Alexander Herbig and Johannes Krause* ✪ *

Abstract | Over the past decade, a genomics revolution, made possible through the development of high-throughput sequencing, has triggered considerable progress in the study of ancient DNA, enabling complete genomes of past organisms to be reconstructed. A newly established branch of this field, ancient pathogen genomics, affords an in-depth view of microbial evolution by providing a molecular fossil record for a number of human-associated pathogens. Recent accomplishments include the confident identification of causative agents from past pandemics, the discovery of microbial lineages that are now extinct, the extrapolation of past emergence events on a chronological scale and the characterization of long-term evolutionary history of microorganisms that remain relevant to public health today. In this Review, we discuss methodological advancements, persistent challenges and novel revelations gained through the study of ancient pathogen genomes.

**Ancient pathogen genomics**
A scientific field focused on the study of whole pathogen genomes retrieved from ancient human, animal or plant remains.

**Neolithic revolution**
The cultural transition associated with the adoption of farming, animal husbandry and domestication as well as the practice of a sedentary lifestyle among human populations.

**Zoonotic transmission**
The infectious disease transmission from animals to humans.

*Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany.*

*\*e-mail: spyrou@shh.mpg.de; krause@shh.mpg.de*

The long shared history between humans and infectious disease places ancient pathogen genomics within the interest of several fields such as microbiology, evolutionary biology, history and anthropology. Research on this topic aims to better understand the interactions between pathogens and their hosts on an evolutionary timescale, to uncover the origins of pathogens and to disentangle the genetic processes involved in their epidemic emergence among human populations. Over the past 10,000 years, major transitions in human subsistence strategies, such as those that accompanied the Neolithic revolution[1], likely exposed our species to a novel range of infectious agents[2]. Closer contact with domesticated animals would have increased the frequency of zoonotic transmission events, and higher human population densities would have enhanced the potential of pathogens to propagate within and between groups. Throughout human history, a number of epidemics and pandemics have been recorded or are hypothesized to have occurred (FIG. 1). Although most of their causative agents still remain speculative, robust molecular methods coupled with archaeological and historical data can confidently demonstrate the involvement of certain pathogens in these episodes.

The investigation of past infectious diseases has traditionally been conducted through palaeopathological assessment of ancient skeletal assemblages[3,4], although this approach is limited by the fact that most acute infections do not leave visible traces on bone. Since the 1990s, the field of ancient DNA (aDNA) has brought molecular techniques to this study, providing a diachronic genetic perspective to infectious disease research. Initial attempts relied on PCR technology[5–9], which restricted the study of ancient microbial DNA to targeted, short genomic fragments that were amplified from ancient human remains. This method made infectious disease detection possible but gave limited information on the evolutionary history of the pathogen. In addition, complications associated with the study of aDNA, which is typically present at low quantities, is heavily fragmented and harbours chemical modifications[10–12], hampered efforts to reproduce and authenticate early findings[13–15].

Over the past decade, major advancements in genomics, in particular, the development of high-throughput sequencing, also called next-generation sequencing (NGS)[16], radically increased the amount of data that can be retrieved from ancient remains. This technology has assisted the development of quantitative methods for aDNA authentication[11,12,17–19] and has enabled the retrieval of whole ancient pathogen genomes from archaeological specimens. The first such genome, published in 2011 (REF.[20]), was that of the notorious bacterial pathogen *Yersinia pestis*, the causative agent of plague. Since then, the field has expanded its directions to the in-depth study of infectious disease evolution, providing a unique resource for understanding human history.

Here, we review the latest methodological innovations that have aided the whole-genome retrieval and evolutionary analysis of various ancient pathogens (TABLE 1), most of which are still relevant to public health today.
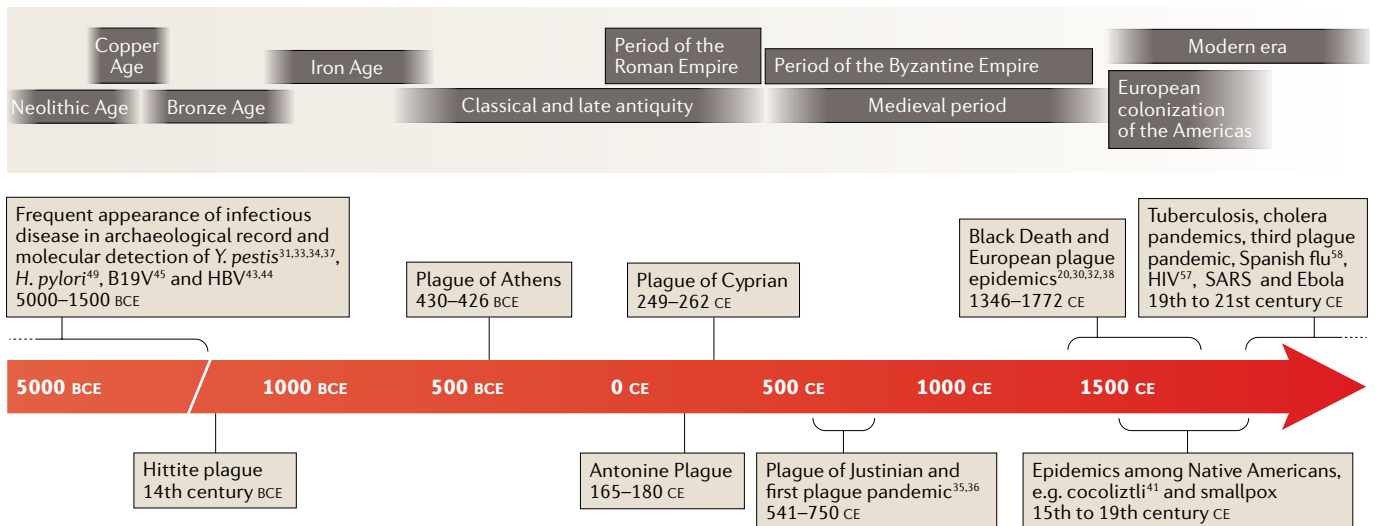
Fig. 1 | **Selected cultural time periods and epidemics or pandemics of human history.** This overview provides a timeline of key events in predominantly Eurasian history since the Neolithic period (upper panel, grey squares), which have overlapped temporally and geographically with major historical epidemics or pandemics (lower panel, beige squares). The respective citations are indicated, in which whole-genome or low-coverage genome-wide data from pathogens implicated in those events have been reconstructed by ancient DNA analysis. B19V, human parvovirus B19; BCE, before current era; CE, current era; HBV, hepatitis B virus; *H. pylori*, *Helicobacter pylori*; SARS, severe acute respiratory syndrome; *Y. pestis*, *Yersinia pestis*.

**Pandemics**
Refers to increased, often sudden, disease occurrence within populations across more than one region or continent, whereas epidemics refers to increased disease occurrences within a confined region or country.

**Palaeopathological assessment**
The evaluation of the health status of ancient individuals or populations, usually through the analysis of disease marker presence on skeletal assemblages.

**Ancient DNA**
(aDNA). The DNA that has been retrieved from historical, archaeological or palaeontological remains.

**Tropism**
Refers to the type of tissue or cell in which infection is established and supported.

**Pulp chamber**
The highly vascularized inner tooth cavity that is contained within the crown and root portions.

**Dental calculus**
Calcified dental plaque.

In the second half of this Review, we highlight the utility of this approach by discussing evolutionary events in the history of *Y. pestis* that have been uniquely revealed through the study of ancient genomes.

## Methods for isolating ancient microbial DNA

*The sweet spot for ancient pathogen DNA.* The retrieval of DNA from ancient human, animal or plant remains carries with it a number of challenges, namely, its limited preservation and hence low abundance, its highly fragmented and damaged state and the pervasive modern-DNA contamination that necessitates a confident evaluation of its authenticity[21,22]. Efficient aDNA recovery is best accomplished via sampling of the anatomical element that contains the highest quantity of DNA from the target organism. For human aDNA analysis, bone and teeth have been the preferred study material, given their abundance in the archaeological record. Recent studies suggest that the inner-ear portion of the petrous bone[23] and the cementum layer of teeth[24] have the greatest potential for successful human DNA retrieval. However, petrous bone sampling and shotgun NGS sequencing of aDNA from five Bronze Age skeletons previously shown to be carrying *Y. pestis* failed to detect the bacterium in this source material, suggesting that its preservation potential for pathogen DNA is low[25].

Direct sampling from skeletal lesions, where present, has proved a rich source of aDNA for some chronic disease-causing bacteria, such as *Mycobacterium tuberculosis*, which was isolated from vertebrae[26]; *Mycobacterium leprae*, which could be isolated from portions of the maxilla and various long bones[27,28]; and *Treponema pallidum* subsp. *pallidum* and *T. pallidum* subsp. *pertenue*, which have been isolated from long bones[29].

Of note, the sampling methods for recovering pathogen DNA do not generally follow a standardized procedure, in part because of the great diversity in tissue tropism and resulting disease progression. In addition, acute blood-borne infections do not typically produce diagnostic bone changes as opposed to those that affect their hosts chronically[3]. Therefore, if infections have caused mortality in the acute phase, as is the case for individuals from epidemic contexts who do not display skeletal evidence of infection, the preferred study material has been the inner cavities of teeth. Pathogen aDNA is thought to be preserved within the remnants of the pulp chamber, likely as part of desiccated blood[8,17]. Consequently, tooth sampling has proved successful in the retrieval of whole genomes or genome-wide data (that is, low-coverage genomes that have provided limited analytical resolution) from ancient bacteria such as *Y. pestis*[20,30–39], *Borrelia recurrentis*[40] and *Salmonella enterica*[41]; ancient eukaryotic pathogens such as *Plasmodium falciparum*[42]; and ancient viruses such as hepatitis B virus (HBV)[43,44] and human parvovirus B19 (B19V)[45]. Even *M. leprae*, which commonly manifests in the chronic form, has been retrieved from ancient teeth[27,28].

Other types of specimen have also shown potential for aDNA retrieval. Examples are dental calculus as a source of oral pathogens, such as *Tannerella forsythia*[46]; calcified nodules, which have yielded whole genomes from *Brucella melitensis*[47], *Staphylococcus saprophyticus* and *Gardnerella vaginalis*[48]; mummified tissues, which have yielded *Helicobacter pylori*[49], Variola virus (VARV; also known as smallpox)[50,51], *M. tuberculosis*[52] and HBV[53,54]; alcohol-preserved human tissue as a source for *Vibrio cholerae* DNA[55]; historical blood stains preserving *P. falciparum* and *Plasmodium vivax*[56]; frozen and formalin-fixed samples, yielding HIV[57] and influenza

Table 1 | **Ancient pathogen genomic data recovered from archaeological or historical specimens**

| Pathogen | Infectious disease | Method of retrieval | Number of genomes[a] | Biological insights | Refs |
|---|---|---|---|---|---|
| **Bacterial pathogens** | | | | | |
| *Borrelia recurrentis* | Relapsing fever | Shotgun sequencing | 1 | • Isolation from 15th-century CE human remains from Norway<br>• Genome signatures of reductive evolution, associated with typical virulence profile, and recent ecological adaptation | 40 |
| *Brucella melitensis* | Brucellosis | Shotgun sequencing | 1 | • Isolation from a calcified nodule identified in an individual's pelvic girdle<br>• Presence of *B. melitensis* in Sardinia during the 14th century CE | 47 |
| *Gardnerella vaginalis* | Bacterial vaginosis | Shotgun sequencing | 1 | • Identified in human remains from Troy dating to 13th century CE<br>• Association with women's mortality during childbirth in the past<br>• The identified strain clusters among modern *G. vaginalis* diversity | 48 |
| *Helicobacter pylori* | • Ulcers of the upper gastrointestinal tract<br>• Increased risk of gastric carcinoma | In-solution capture followed by NGS | 1 | • Isolation from European Copper Age, 5,300-year-old mummy (Ötzi)<br>• Unadmixed strain, contrary to modern European strains, which are hybrids of two ancestral populations | 49 |
| *Mycobacterium leprae* | Lepromatous leprosy | • Shotgun sequencing<br>• Microarray-based capture followed by NGS | 27 | • First de novo assembled ancient pathogen genome<br>• Estimated emergence >5,000 years ago<br>• European origin of leprosy in the Americas<br>• High *M. leprae* diversity in medieval Europe | 27,28, 105,201 |
| *Mycobacterium tuberculosis* | Tuberculosis | • Shotgun sequencing<br>• Microarray-based capture followed by NGS | 19 | • Genomes from pre-Columbian human infections show phylogenetic clustering within animal-adapted lineage present today in seals<br>• Molecular dating analysis suggests emergence of MTBC <6,000 years ago<br>• Analysis of European genomes shows past occurrence of multiple infections and suggests origin of lineage 4 during the 4th to 5th century CE | 26,52,123 |
| *Salmonella enterica* subsp. *enterica* serovar Paratyphi C | Enteric (paratyphoid) fever | • Shotgun sequencing<br>• Microarray-based capture followed by NGS<br>• In-solution capture followed by NGS | 11 | • *S. enterica* subsp. *enterica* serovar Paratyphi C presence in 12th-century CE Norway<br>• Paratyphi C serovar was also identified among 16th-century individuals from Mexico that were associated with the major post-contact 'cocoliztli' epidemic | 41,109 |
| *Staphylococcus saprophyticus* | • Urinary tract infections<br>• Puerperal fever | Shotgun sequencing | 1 | • Identified in ~800-year-old human remains from Troy<br>• Association with women's mortality during childbirth in the past<br>• The identified lineage is not commonly associated with human disease today | 48 |
| *Tannerella forsythia* | Periodontal disease | Shotgun sequencing | 1 | • Isolation from medieval human remains (circa 950–1200 CE)<br>• First pathogen genome reconstructed from ancient dental calculus | 46 |
| *Treponema pallidum* | • Syphilis (*Treponema pallidum* subsp. *pallidum*)<br>• Yaws (*Treponema pallidum* subsp. *pertenue*)<br>• Bejel (*Treponema pallidum* subsp. *endemicum*) | Microarray-based capture followed by NGS | 3 | • Isolated from individuals who lived in Mexico City between the 17th and 19th centuries CE<br>• Different Treponema subspecies (*T. pallidum* subsp. *pallidum* and subsp. *pertenue*) caused similar skeletal lesions usually identifiable as skeletal syphilis in infants | 29 |
| *Vibrio cholerae* | Cholera | Microarray-based capture followed by NGS | 1 | • Isolation from 19th-century alcohol-preserved intestinal specimen from an individual affected during the second cholera pandemic<br>• The identified strain shows highest similarity with the classic pathogenic biotype O1 | 55 |
| *Yersinia pestis* | Bubonic, pneumonic and septicaemic plague | • Shotgun sequencing<br>• Microarray-based capture followed by NGS<br>• In-solution capture followed by NGS | 38 | • Bacterium affected humans as early as 5,000 years ago<br>• Both flea-adapted and non-adapted variants were present in Eurasia during the Bronze Age<br>• Causative agent of the Plague of Justinian (6th century CE)<br>• Causative agent of Black Death and persistence in Europe during the second plague pandemic (14th to 18th century CE)<br>• Possible European origin of third plague pandemic lineage | 20,30–39 |

Table 1 (cont.) | Ancient pathogen genomic data recovered from archaeological or historical specimens

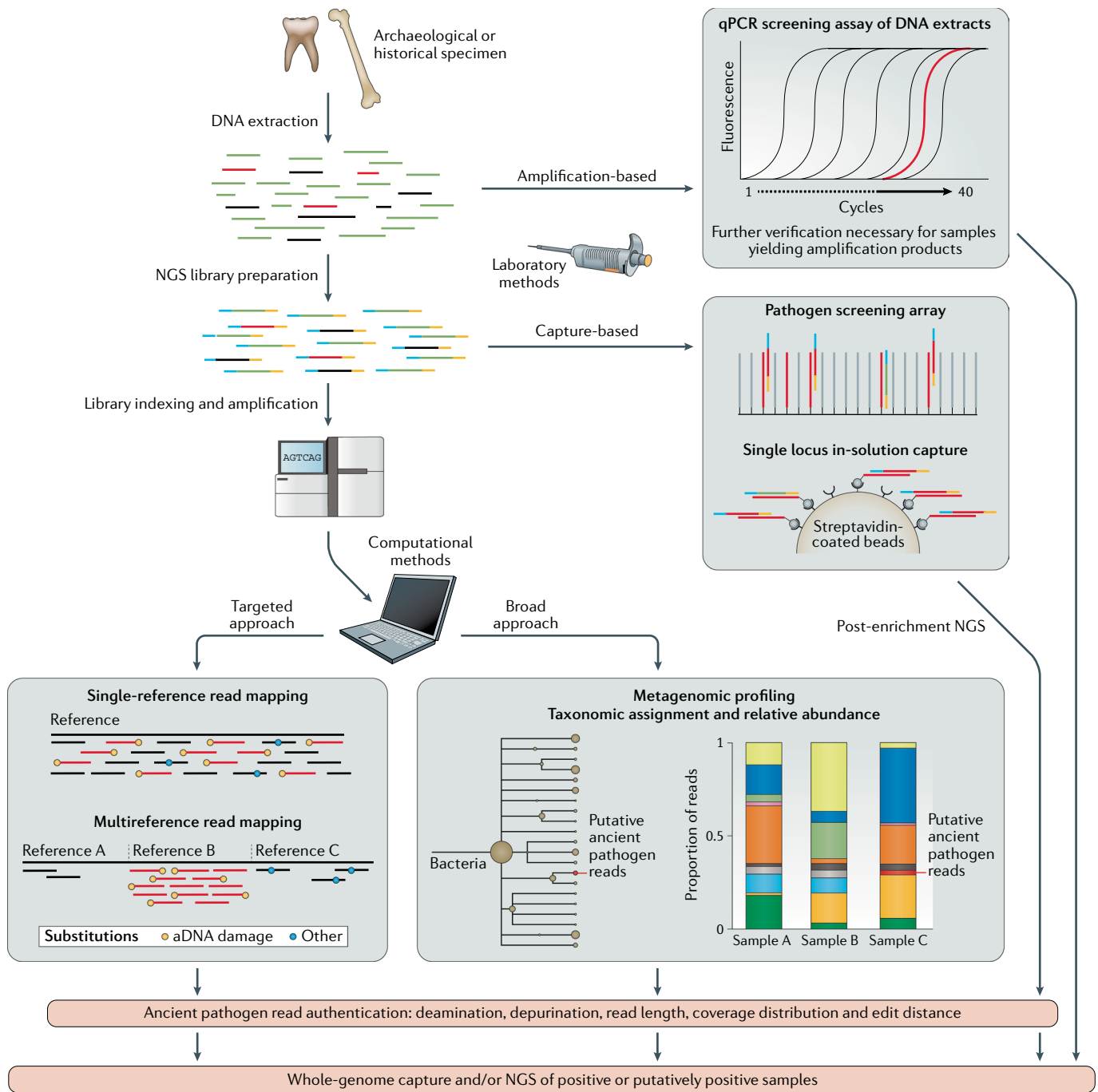| Pathogen | Infectious disease | Method of retrieval | Number of genomes[a] | Biological insights | Refs |
|---|---|---|---|---|---|
| **Viral pathogens** | | | | | |
| HBV | Viral hepatitis | • Shotgun sequencing<br>• In-solution capture followed by NGS<br>• Whole-genome PCR[b] | 17 | • Identified in ancient human specimens as early as 7,000 years ago<br>• Neolithic genome lineage related to contemporary strains identified in African non-human primates<br>• Complex evolutionary history of HBV and identification of ancient recombination event giving rise to genotype A strains | 43,44,53,54 |
| HIV | AIDS | Whole-genome PCR[b] | 8 | • Analysis of HIV RNA from archival specimens of seropositive individuals enrolled in HBV studies during the late 1970s<br>• HIV was introduced into the Americas from the Caribbean in the early 1970s | 57 |
| B19V | • Erythema infectiosum (fifth disease) in children<br>• Arthropathies in adults<br>• Hydrops fetalis or fetal death in pregnant women<br>• Pure red-cell aplasia | In-solution capture followed by NGS | 10 | • Genomic signatures of B19V identified in human remains dating as early as ~7,000 years ago<br>• Contrary to previous estimates of a most recent common ancestor younger than 200 years, phylogenetic and molecular dating analysis of ancient genomes showed a much lengthier association of B19V with human populations | 45 |
| Influenza virus | Influenza | Whole-genome PCR[b] | 1 | • First reconstructed genome from historical RNA virus<br>• Avian source of 1918 influenza pandemic (Spanish flu, 1918–1920)<br>• Reconstructed virus particle displayed increased virulence under laboratory conditions | 58,202 |
| VARV | Smallpox | In-solution capture followed by NGS | 1 | • Genome reconstruction from a 17th-century mummy from Lithuania<br>• Recent emergence of 20th century VARV lineages (divergence during the 18th century CE) | 50 |
| **Eukaryotic pathogens** | | | | | |
| *Phytophthora infestans* | Late blight (also known as potato blight) | Shotgun sequencing | 18 | • First sequenced ancient eukaryotic (plant) pathogen genomes<br>• Isolated from historical herbarium specimens<br>• A unique *Phytophthora infestans* genotype caused the Irish potato famine and during the 1900s became replaced by the US-1 lineage that dominated worldwide until the 1970s | 59,60 |
| *Plasmodium falciparum* and *Plasmodium vivax* | Malaria | In-solution capture followed by NGS | 5 | • Oldest *Plasmodium falciparum* detection from southern Italy (1st to 2nd century CE)<br>• *Plasmodium falciparum* and *Plasmodium vivax* mitochondrial genome isolation from 20th century microscopy slides<br>• Possible introduction of *Plasmodium vivax* in the Americas through European contact | 42,56 |

B19V, human parvovirus B19; CE, current era; HBV, hepatitis B virus; MTBC, *Mycobacterium tuberculosis* complex; NGS, next-generation sequencing; VARV, variola virus. [a]The indicated numbers include whole pathogen genomes and specimens yielding genome-wide data. [b]Whole-genome PCR amplicons from the studies of influenza virus[58], HIV[57] and HBV[54] that were sequenced using capillary sequencing (Sanger method).

virus[58] RNA; and dried plant leaves from herbarium collections, preserving *Phytophthora infestans*[59,60], the oomycete that caused the Irish potato famine.

**Segregating the metagenomic soup: methods for pathogen detection.** Regardless of the source of genetic material, most ancient specimens yield complex metagenomic data sets. Poorly preserved aDNA usually makes up a miniscule fraction of the total genetic material extracted from a sample (<1%), and the majority of DNA usually stems from organisms residing in the environment[41]. Hence, specialized protocols are necessary for the detection and isolation of ancient pathogen DNA and its confident segregation from a rich environmental DNA background (FIG. 2).

**Metagenomic**
A term used to describe a specimen or data set that includes nucleic acid sequences from all organisms within the sampled proportion.

In this context, laboratory-based techniques are separated into those that target a specific microorganism and those that screen for several pathogenic microorganisms simultaneously (FIG. 2). Methods that screen for a single microorganism have used species-specific assays of conventional or quantitative PCR (also known as real-time PCR)[17,61–64], as well as hybridization-based enrichment techniques[17,26,28] (FIG. 2). These methods are particularly useful when the target microorganism is known, for example, in the presence of diagnostic skeletal lesions among the studied individuals[26,28], or when a hypothesis exists for the causative agent of an epidemic[17]. By contrast, broad laboratory-based pathogen screening in aDNA research has used microarrays for both targeted enrichment[65] and fluorescence-based

Fig. 2 | **Methods for the detection and isolation of pathogen DNA from ancient metagenomic specimens.** The diagram provides an overview of techniques used for pathogen DNA detection in ancient remains by distinguishing between laboratory and computational methods. In both cases, processing begins with the extraction of DNA from ancient specimens[183]. As part of the laboratory pipeline, direct screening of extracts can be performed by PCR (quantitative (qPCR) or conventional) against species-specific genes, as done previously[17,61,63,64]. PCR techniques alone, however, can suffer from frequent false-positive results and should therefore always be coupled with further verification methods such as downstream genome enrichment and/or next-generation sequencing (NGS) in order to ensure ancient DNA (aDNA) authentication of putatively positive samples. Alternatively, construction of NGS libraries[184,185] has enabled pathogen screening via fluorescence-based detection on microarrays[66] and via DNA enrichment approaches[17]. The latter has been achieved, through single locus in-solution capture[26,28] or through simultaneous screening for multiple pathogens using microarray-based enrichment of species-specific loci[65] and enables post-NGS aDNA authentication. In addition, data produced by direct (shotgun) sequencing of NGS libraries before enrichment can also be used for pathogen screening using computational tools. After pre-processing, reads can be directly mapped against a target reference genome (in cases for which contextual information is suggestive of a causative organism) or against a multigenome reference composed of closely related species to achieve increased mapping specificity of ancient reads. Alternatively, ancient pathogen DNA can also be detected using metagenomic profiling methods, as presented elsewhere[41,71,72], through taxonomic assignment of shotgun NGS reads. Both approaches allow for subsequent assessment of aDNA authenticity and can be followed by whole pathogen genome retrieval through targeted enrichment or direct sequencing of positive sample libraries.

detection[66], whereby probes are designed to represent unique or conserved regions from a range of pathogenic bacteria, parasites or viruses. Although amplification-based or fluorescence-based approaches can be fast and cost-effective for screening large sample collections[17,38], enrichment-based techniques are usually coupled with NGS and therefore provide data that can be used to assess aDNA authenticity.

When shotgun-sequencing data are generated, computational screening approaches can be used to detect the presence of pathogen DNA as well as for metagenomic profiling of ancient specimens (FIG. 2). In cases for which a causative agent is suspected, NGS reads can be directly mapped (for example, using the read alignment software Burrows–Wheeler aligner[67]) against a specific reference genome or against a multigenome reference that includes several species of a certain genus with the purpose of achieving a higher mapping specificity to the target organism[34] (FIG. 2). In addition, broad approaches involve the use of metagenomic techniques for pathogen screening. Examples of tools that have shown their effectiveness with ancient metagenomic DNA include the widely used Basic Local Alignment Search Tool (BLAST)[68]; the MEGAN Alignment Tool (MALT)[41], which involves a taxonomic binning algorithm that can use whole-genome databases (such as the National Center for Biotechnical Information (NCBI) Reference Sequence (RefSeq) database[69]); Metagenomic Phylogenetic Analysis (MetaPhlAn)[70], which is also integrated into the metagenomic pipeline MetaBIT[71] and uses thousands (or millions) of marker genes for the distinction of specific microbial clades; or Kraken[72], an alignment-free sequence classifier that is based on k-mer matching of a query to a constructed database.

Taxonomic sequence assignments from the above methods, however, should be interpreted with caution, mainly because some pathogenic microorganisms have close environmental relatives that are often insufficiently represented in public databases. For example, a >97% sequence identity was shown between environmental taxa and human-associated pathogens such as *M. tuberculosis* and *Y. pestis* according to an analysis of 16S ribosomal RNA genes[73]. As such, given that environmental DNA often dominates ancient remains that stem from burial contexts[74], analyses should always ensure a qualitative assessment of assigned reads, that is, an evaluation of their mapping specificity and their genetic distance (also called edit distance) to the putatively detected organism. In addition, one should consider the known aDNA damage characteristics as criteria for data authenticity. Although several types of chemical damage can affect post-mortem DNA survival, certain characteristics have been more extensively quantified. The first, termed depuration, is a hydrolytic mechanism under which purine bases become excised from DNA strands. This process results in the formation of abasic sites and is a known contributor to the fragmentation patterns observed in aDNA. As such, an increased base frequency of A and G compared with C and T immediately preceding the 5′ ends of aDNA fragments is often considered a criterion for authenticity[12]. A second type of damage commonly identified among aDNA

data sets is the hydrolytic deamination of C, whereby a C base is converted into U (and detected as its DNA analogue, T)[12,75]. This base modification usually occurs at single-stranded DNA overhangs that are most accessible to environmental insults, resulting in an increased frequency of miscoding lesions at the terminal ends of aDNA fragments[11,12]. Consequently, the evaluation of DNA damage profiles (for instance, by using map-Damage2.0 (REF.[76])) is a prerequisite for authenticating ancient pathogen DNA and is necessary for ensuring aDNA data integrity in general. More detailed overviews of authentication criteria in ancient pathogen research have been reviewed elsewhere[19,73].

***Targeted enrichment approaches to isolate whole ancient pathogen genomes.*** Evolutionary relationships between past and present infectious agents are best determined through the use of whole-genome sequences of pathogens. However, the recovery of high-quality data is often challenging owing to the aforementioned characteristics of aDNA and therefore requires specialized sample processing. For example, in cases in which aDNA authenticity has already been achieved in the detection step, U residues resulting from post-mortem C deamination can be entirely[77] or partially[78] excised from aDNA molecules using the enzyme uracil-DNA glycosylase (UDG) to avoid their interference with downstream read mapping and variant calling.

In addition, given the low proportion of pathogen DNA in ancient remains, a common and cost-effective approach for whole-genome retrieval involves microarray-based or in-solution-based hybridization capture. Both methods constitute a form of genomic selection of continuous or discontinuous genomic regions through the design and use of single-stranded DNA or RNA probes that are complementary to the desired target. Microarray-based capture utilizes densely packed probes that are immobilized on a glass slide[79]. It is cost-effective in that it permits the parallel enrichment of molecules from several libraries that can be subsequently recovered through deep sequencing, although competition over the probes can impair enrichment efficiencies in specimens with comparatively lower target DNA contents. Nevertheless, this type of capture has shown its effectiveness in the recovery of both ancient pathogen and human DNA[20,26,28,41,55,80].

More recently, in-solution-based capture approaches have gained popularity owing to their capacity for greater sample throughput without compromising capture efficiency[81–83]; every sample library can be captured individually, thus providing, in principle, an equal probe density per specimen. This technique has contributed to the increased number of specimens from which human genome-wide single-nucleotide polymorphism (SNP) data could be retrieved[84,85], even from climate zones that pose challenges to aDNA preservation (presented elsewhere[86–88]). In addition, in-solution-based capture has recently become the preferred method for microbial pathogen genome recovery for both bacteria and DNA viruses (for examples, see REFS[34,37,41,43,45,49,50]). Nevertheless, deep shotgun sequencing alone has also been used for human[89–91] and pathogen[28,33,48] high-quality

**Taxonomic binning**
An algorithm that assigns metagenomic DNA reads to a species or a higher taxonomic rank (for example, genus or family) based on the sequence specificity.

**k-mer matching**
The matching, for each read, of multiple subsequences of length k without mismatches to a database.

**Depuration**
A hydrolytic reaction in which the β-*N*-glycosidic bond of a purine (adenine or guanine) is cleaved, causing its excision from a DNA strand.

**Deamination**
The hydrolytic removal of an amine group (NH₂) from a molecule. In ancient DNA studies, the term deamination most often refers to the deamination of cytosine residues into uracils.

**Variant calling**
The identification of polymorphisms (nucleotide differences) in sequenced data by comparison to a reference.

genome reconstruction, especially for specimens with fairly high endogenous DNA yields, although this frequently carries with it a greater production cost.

### Disentangling microbial evolution

***Ancient pathogen genomes as molecular fossils.*** In the absence of ancient pathogen genomes, the timings of infectious disease emergence and early spread are inferred mainly through comparative genomics of modern pathogen diversity[92,93], palaeopathological evaluation of ancient skeletal remains[94] or analysis of historical records[95,96]. Such approaches are highly valuable and, when combined, can be used to build an interdisciplinary picture of infectious disease history; however, limitations also exist. For example, the analysis of contemporary pathogen genetic diversity considers only a short time depth of available data and cannot predict evolutionary scenarios that derive from lineages that are now extinct. In addition, skeletal markers of specific infections in past populations only exist for a few conditions and, when present, can rarely be considered as definitive, as numerous differential diagnoses can exist for a given skeletal pathology[97]. Similarly, historically recorded symptoms can often be misinterpreted given that past descriptions may be unspecific and do not always conform to modern medical terminology[98].

In the past decade, the reconstruction of ancient pathogen genomes has complemented such analyses with direct molecular evidence, often revealing aspects of past infections that were unexpected on the basis of existing data. The recent identification of HBV DNA in a mummified individual showing a vesicopustular rash[53], which is usually considered characteristic of infection with VARV, highlights the importance of molecular methods in evaluating differential diagnoses. The oldest recovered genomic evidence of HBV to date was from a 7,000-year-old individual from present-day Germany[44], which shows that this pathogen has affected human populations since the Neolithic period. In addition, the virus was identified recently in human remains from the Bronze Age, Iron Age and up until the 16th century of the current era (CE) in Eurasia[43,44,53,54].

Regarding bacterial pathogens, the identification of *B. recurrentis* in a 15th-century individual from Norway[40] showed that — aside from *Y. pestis* — other vector-borne pathogens were also circulating in medieval Europe. Furthermore, the causative agents of syphilis and yaws, *T. pallidum* subsp. *pallidum* and *T. pallidum* subsp. *pertenue*, respectively, were recently identified in different individuals from colonial Mexico[29] who exhibited similar skeletal lesions. This study demonstrates the power of ancient pathogen genomics in distinguishing past infectious disease agents that are genetically and phenotypically similar but that differ greatly in their public health significance. Finally, the identification of *G. vaginalis* and *S. saprophyticus* in calcified nodules from a woman's remains (13th-century Troy)[48] directly implicates these bacteria in pregnancy-related complications in the past. These findings, as well as other insights gained from analyses of ancient pathogen genomes (TABLE 1), demonstrate the ability of aDNA to contribute aspects of infectious disease history beyond those accessible by the palaeopathological, historical and modern genetic records.

***Assessing within-species evolutionary relationships.*** The reconstruction of whole pathogen genomes has not only been a tool for demonstrating infectious disease presence in the past but also aided in the robust inference of microbial phylogeography, which is important for understanding the processes that influence pathogen distribution and diversity over time.

The evaluation of genetic relationships between ancient and modern pathogens is often conducted by direct whole-genome or genome-wide SNP comparisons of bacteria[20,27,29,36,48], viruses[43,44,50,53] or mitochondrial genomes and nuclear genome data from eukaryotic microorganisms[56,59,60]. Hence, accurate variant calling is critical for drawing reliable evolutionary inferences, although this process is often a challenge when handling data sets derived from samples with high rates of DNA fragmentation (resulting in ultrashort read data), low endogenous DNA content and high levels of DNA damage. In these cases, increased accuracy is best achieved through stringent NGS read mapping parameters and through visual inspection of the sequences overlapping the studied SNPs[35]. In addition, histograms of SNP allele frequencies — used to estimate the frequency of heterozygous calls in haploid organisms[26,52] — can often demonstrate the effects of environmental contamination on ancient microbial data sets[41].

Once variant calls are authenticated, one of the most common types of evolutionary inference in pathogen research is through phylogenetic analysis, which is a powerful means of resolving the genetic history of clonal microorganisms (FIG. 3). Among the most commonly used tools in ancient microbial genomics are MEGA[99], which comprises several phylogenetic methods; PhyML[100], RAxML[101] and IQ-TREE[102], which implement maximum-likelihood approaches; MrBayes[103], which uses a Bayesian approach; and programs used for phylogenetic network inference, such as SplitsTree[104]. Two notable studies that examined phylogenetic relationships among ancient *M. leprae* genomes revealed a high strain diversity in Europe between the 5th and 14th centuries CE[27,105]. Considered alongside the oldest palaeopathological cases of leprosy dating to as early as the Copper and Bronze Age in Eurasia[106,107] and the high frequency of protective immune variants against the disease identified in modern-day Europeans[108], these results may suggest a long history of *M. leprae* presence in this region. Moreover, the phylogenetic analysis of a 12th-century *S. enterica* subsp. *enterica* genome from Europe showed its placement within the Paratyphi C lineage[109]. Further identification of the bacterium in 16th-century colonial Mexico[41] revealed it as a previously unknown candidate pathogen that was likely introduced to the Americas through European contact. Given the low frequency of Paratyphi C today, these results may be indicative of a higher prevalence in past populations. Finally, an example from viral genomics is the recovery of HIV RNA from degraded serum specimens[57], which highlighted the importance of archival collections in reconciling the expansion of recent pandemics. Specifically, these
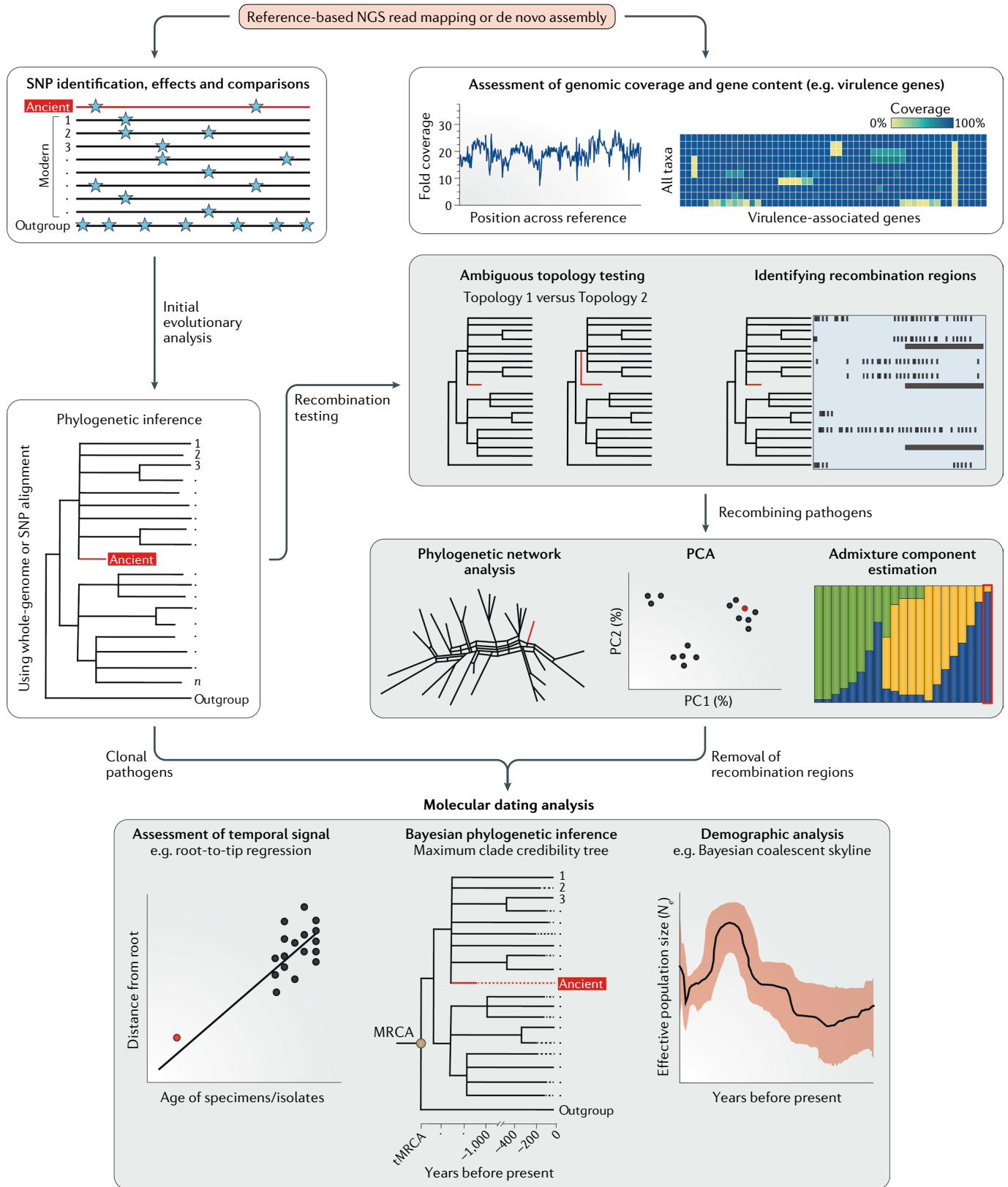
data were able to dispute a long-standing hypothesis regarding the initiation of HIV spread in the USA.

When the evolutionary histories of pathogens are influenced equally by mutation and recombination, additional tools have been used to identify recombining loci and to determine genetic relationships within and between microbial populations (FIG. 3). For example, the programs ClonalFrameML[110] and Recombination Detection Program 4 (RDP4)[111] have been used to infer potential recombination regions within ancient

◄ Fig. 3 | **Methods for whole-genome analysis of clonal and recombining pathogens.** The diagram is an overview of whole-genome analysis applied to date for ancient microbial data sets and distinguishes the methods used for clonal and recombining pathogens; of note, the depicted summary is not meant to represent an exhaustive pipeline of all possible analyses that could be undertaken. Ancient genome reconstruction is usually initiated through reference-based mapping or through de novo assembly of the data, although the latter has only been possible in exceptional cases of ancient DNA (aDNA) preservation[28,44]. Subsequently, the genomes are assessed for their coverage depth and gene content for evaluation of their quality, which is also relevant for the comparative identification of virulence genes over their evolutionary time frames. Here, we show an example of virulence factor presence-or-absence analysis in the form of a heat map, as done previously[33,34,37,41]. In addition, a comparison of the ancient genome or genomes with modern genomes can be carried out for single-nucleotide polymorphism (SNP) identification and for assessment of SNP effects (using SnpEff[186]), which is particularly relevant for variants that seem to be unique to the ancient genome or genomes. Initial evolutionary inference can often be carried out through phylogenetic analysis and by testing for possible evidence of recombination in the analysed data set, for example, by comparing the support of different phylogenetic topologies[114] and by identifying potential recombination regions and homoplasies[110,111]. If the data support clonal evolution, robust phylogenetic inference (for example, through a maximum-likelihood approach) is followed by assessment of the temporal signal in the data[124,125]. If the data set shows a sufficient phylogenetic signal, molecular dating analysis and demographic modelling are considered possible, although the size of the data set will determine whether such analyses will be feasible and meaningful. Alternatively, if recombination is confirmed, genetic relationships between microbial clades or populations can be determined through phylogenetic network analysis[104] or through the use of population genetic methods such as principal component analysis (PCA) and identification of ancestral admixture components[112,113]. In this case, the assessment of the temporal signal and proceeding with molecular dating analysis is cautioned and likely best performed after exclusion of recombination regions from all genomes in the data set. MRCA, most recent common ancestor. NGS, next-generation sequencing.

**Molecular clock**
A term used to describe that genome evolution occurs as a function of time and, therefore, the genetic distance between two living forms is proportional to the time of their divergence.

**Radiocarbon dating**
A technique to estimate the age of a specimen on the basis of the amount of incorporated radiocarbon (^{14}C) that after the death of an organism gradually becomes lost over time.

**Nucleotide substitution rate**
Denotes the frequency of substitution accumulation in an organism within a given time; usually represented as substitutions per site per year.

**Divergence dates**
The dates of separation between two phylogenetic lineages, for example, the split between two species.

bacteria[29,37] and viruses[43,44,53], respectively. In addition, principal component analysis (PCA) and ancient admixture component estimation using the Bayesian modelling frameworks STRUCTURE[112] and fineSTRUCTURE[113] on both multilocus sequence typing (MLST) and whole-genome data were recently used for population assignment of a 5,300-year-old *H. pylori* genome[49]. These analyses revealed key information on changes of the bacterial population structure that occurred in Europe over time. Furthermore, the recent study of ancient *T. pallidum* subsp. *pallidum* and *T. pallidum* subsp. *pertenue*[29] used the program TREE-PUZZLE[114], a maximum-likelihood-based phylogenetic algorithm, to gain a more robust phylogenetic resolution of ambiguous branching patterns among bacterial lineages.

Such whole-genome analyses of both clonal and recombining pathogens have helped to elucidate not only past infectious disease phylogeography but also possible zoonotic or anthroponotic transmission events that reveal disease interaction networks through time. Among others (TABLE 1), a notable example is that of 1,000-year-old pre-Columbian *M. tuberculosis* genomes isolated from human remains, which showed a phylogenetic placement among animal-adapted lineages, being most closely related to a strain circulating in modern-day seals and sea lions[26]. Although the extent to which these strains were capable of human-to-human transmission is unclear, this study supports the existence of tuberculosis in pre-Columbian South America and is helping to delineate the genomic and adaptive history of *M. tuberculosis* in the region before European contact[26]. Another example of intriguing

evolutionary relationships revealed uniquely through the study of ancient pathogen genomes includes analyses of Neolithic and Bronze Age HBV. These genomes grouped in extinct lineages that are most closely related to modern strains identified exclusively among African non-human primates[43,44], a result that raises further questions regarding past transmission events in HBV history. Finally, the phylogenetic analysis of medieval *M. leprae* genomes suggested a European source for leprosy in the Americas[28], reinforcing the hypothesis that humans passed the disease to the nine-banded armadillo, the most common reservoir for this disease in the New World[115].

Importantly, the resolution of evolutionary analyses will depend on the quality, size and evenness of spatial sampling in the comparative data set. Therefore, the incomplete and often biased sampling of ancient and modern microbial strains can introduce challenges for discerning true biological relationships and past evolutionary events. Nevertheless, in recent years, marked reductions in NGS costs[116] have aided the increased production of large whole-genome microbial data sets from present-day strains. Current efforts for centralized data repositories that are continuously curated (such as the Pathosystems Resource Integration Center (PATRIC) database[117] and the recently introduced EnteroBase[118]) and the development of robust phylogenetic frameworks that can accommodate genome-wide data from >100,000 strains (for example, GrapeTree[119]) are becoming valuable for integrating large sample sizes into microbial evolutionary analyses. In combination with the increasing number of ancient microbial data sets, these tools will aid in the evaluation of genetic relationships by offering higher resolution.

***Inferring divergence times through molecular dating.*** Apart from providing a molecular fossil record and revealing diachronic evolutionary relationships, a third analytical advantage gained from the retrieval of ancient pathogen genomes is that their ages can be directly used for calibration of a molecular clock. The ages of ancient specimens can be determined through contextual information, through archaeological artefacts or directly through radiocarbon dating, predominantly of bone or tooth collagen. Such temporal calibrations are required for high-accuracy estimations of microbial nucleotide substitution rates and in turn lineage divergence dates (FIG. 3), particularly because both estimations seem to be highly influenced by the time depth covered by the genomic data set[120]. For such analyses, the most widely used program is the Bayesian statistical framework BEAST[121,122].

A characteristic example of how ancient calibration points can considerably affect divergence date estimates is that of *M. tuberculosis*. According to modern genetic data and human demographic events, the *M. tuberculosis* complex (MTBC) evolution was suggested to have followed human migrations out of Africa, with its emergence estimated at more than 70,000 years ago[93]. Recently, its emergence was re-estimated to a maximum of 6,000 years ago on the basis of the 1,000-year-old mycobacterial genomes from Peru[26], a result that was further

corroborated by the incorporation of 18th-century European MTBC genomes in the dating analysis[26,52,123].

In molecular phylogenies, the length of each individual branch usually reflects the number of substitutions acquired by an organism within a given period of time and, as such, varying branch lengths should represent heterochronous sequences. Therefore, an important prerequisite for a robust dating analysis is that the nucleotide substitution rate of the species whose phylogeny is to be dated behaves in a 'clock-like' manner, meaning that phylogenetic branch lengths correlate with archaeological dates or sampling times. Such relationships can be assessed through date randomization and root-to-tip regression tests (FIG. 3). The former is used to assess the effect of arbitrary exchange of phylogenetic tip dates on the nucleotide substitution rate and divergence date estimates[124], whereas the latter is used for estimation of a correlation coefficient ($r$) and coefficient of determination ($R^2$) by relating the tip date of each taxon to its SNP distance from the tree root (using, for example, the program TempEst[125]). The resulting values determine whether there is a temporal signal in the data and suggest whether branches within a phylogeny evolve at a constant rate, in which case a strict molecular clock[126] can be statistically tested, for example, using MEGA[99] or marginal likelihood estimations[127,128], and applied. If branches are affected by differences in their evolutionary rates, a relaxed clock[129] would be more appropriate. In general, a constant molecular clock will rarely reliably describe the history of a microbial species, even more so for infectious pathogens whose replication rates vary between active and latent or between epidemic and dormant phases[120,130]. In certain cases, neither of the two models may fit the data, such as when extensive rate variation weakens the temporal signal. This challenge was encountered in initial attempts to date the Y. pestis phylogeny using too few ancient calibration points[36,130]. Similar limitations can arise when the evolutionary history of a microorganism is vastly affected by recombination, as observed for HBV[44,53], although HBV molecular dating was recently attempted using a different genomic data set and suggested that the currently explored diversity of Old and New World primate lineages (including all human genotypes) may have emerged within the last 20,000 years[43].

Molecular dating analysis requires the use of an appropriate demographic model for the available data, which can be determined through model-testing approaches (for example, through marginal likelihood estimations[127,128]). Currently, the most widely used models for estimating dates of divergence are the coalescent constant size[131], which assumes a continuous population size history — and is unrealistic for epidemic pathogens — and the coalescent skyline[132], which can estimate effective population size ($N_e$) changes over time. Moreover, the birth–death demographic model[133,134], which is currently unexplored within aDNA frameworks, may prove an insightful analysis tool in the future. This model has shown its applicability on comprehensive pathogen data sets from modern-day epidemic contexts[133]. It has the ability to incorporate prior knowledge on incomplete sampling proportions and sampling biases within a data set, a frequent caveat of aDNA studies that is currently unaccounted for within molecular dating analyses. Finally, recently developed fast-dating algorithms should also be noted, for example, the least-squared dating (LSD) program, which does not use constrained demographic models but can handle uncorrelated rate variation among phylogenetic branches and has shown potential for analysing large genomic data sets[135].

## Yersinia pestis evolution

The pathogen best studied using aDNA analysis so far is Y. pestis, the causative agent of plague. To date, 38 ancient genomes of this bacterium have been published[20,30–39] (FIG. 4), and their analyses have yielded valuable information on past pandemic emergence as well as in-depth microbial evolution. Integration of such knowledge into human population frameworks has provided key insights into the association of human migrations and infectious disease transmission in the past[31,34]. This section describes the evolutionary history of Y. pestis with the aim of demonstrating aspects of its emergence and spread as revealed through aDNA research.

*Not a human pathogen: plague ecology.* Plague is a well-defined infectious disease caused by the Gram-negative bacterium Y. pestis, which belongs to the family Enterobacteriaceae. It evolved from a close relative, *Yersinia pseudotuberculosis*, which is an environmental enteric-disease-causing bacterium[136]. Although the two species are clearly distinguishable in terms of their virulence potential and transmission mechanisms, their nucleotide genomic identity reaches 97% among chromosomal protein-coding genes[137]. In addition, they share the virulence plasmid pCD1, which encodes a type III secretion system common to three known pathogenic *Yersinia*: Y. pestis, Y. pseudotuberculosis and *Yersinia enterocolitica*. The distinct transmission mechanism and pathogenicity of Y. pestis are conferred by the unique acquisition of two plasmids, pPCP1, which contributes to the invasive potential of the bacterium[138], and pMT1, which is involved in flea colonization[139,140], as well as by chromosomal gene pseudogenization or loss throughout its evolutionary history[141].

Y. pestis is not human adapted. Its primary hosts are sylvatic rodents such as marmots, mice, great gerbils, voles and prairie dogs, among others, in which it is continuously or intermittently maintained in so-called reservoirs or foci[142–144]. Its global distribution includes numerous rodent species[144,145] and encompasses regions in eastern Europe, Asia, Africa and the Americas (FIG. 4), where the bacterium persists in active foci, some of which have existed for centuries or even millennia[31,33,34,37,130]. Y. pestis transmission among hosts is facilitated by a flea vector (FIG. 5). The best yet characterized is the oriental rat flea, *Xenopsylla cheopis*, although others are also known to play important roles in Y. pestis transmission[142,144,146]. Notably, recent modelling inferences suggest important roles for ectoparasites such as body lice and human fleas in its propagation during human epidemics[147]. Landmark studies investigating the classical model of transmission have shown that Y. pestis has the unique ability to colonize and form a biofilm within the flea, which blocks a portion of its
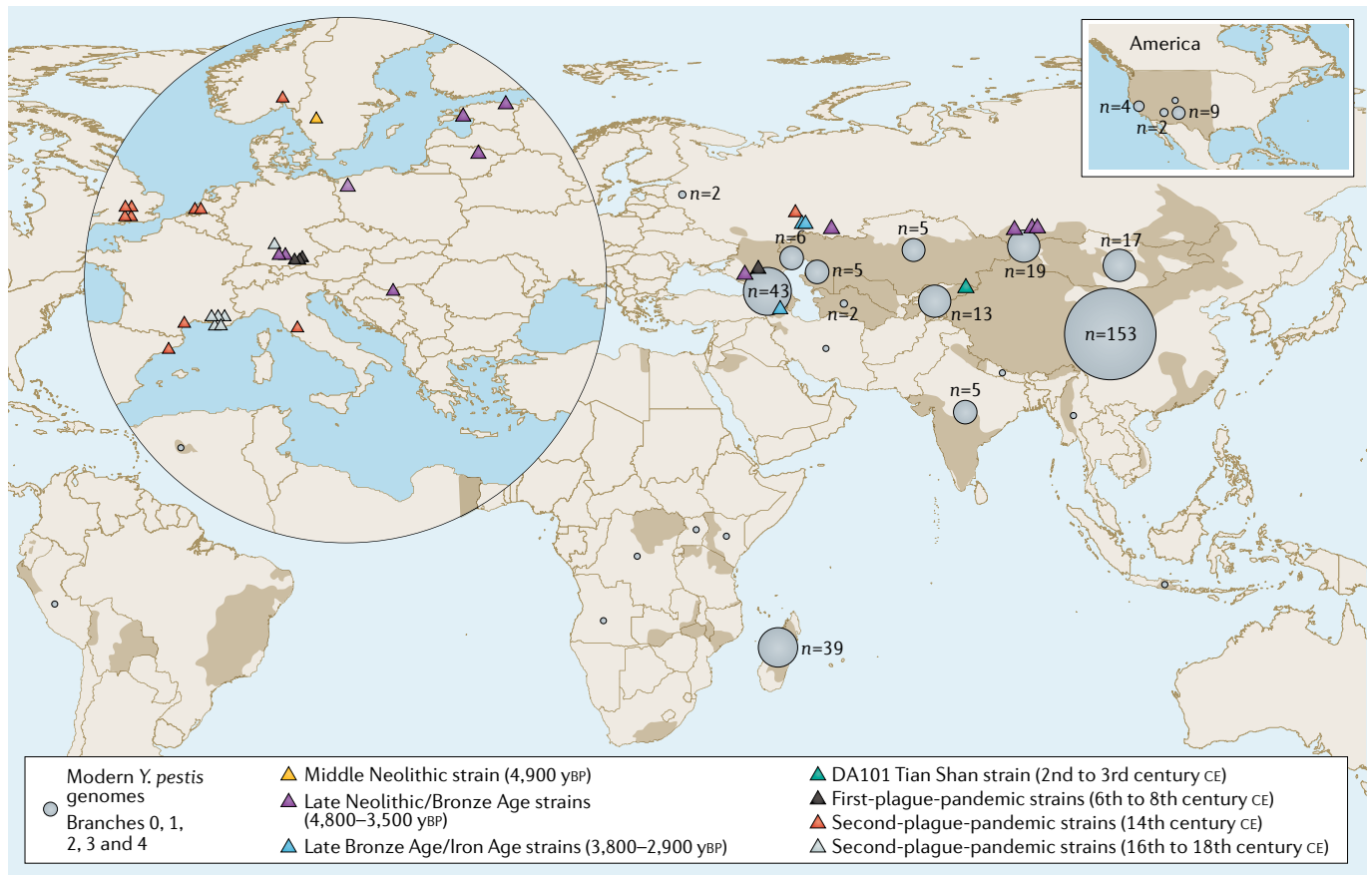
Fig. 4 | **Map of published modern and ancient *Yersinia pestis* genomes.** Published ancient specimens that have yielded whole *Yersinia pestis* genomes and genome-wide data are shown in triangles ($n = 38$), and their different colours indicate time period distinctions. A set of modern *Y. pestis* genomes ($n = 336$), from the following publications (released until 2018)[92,130,169–173,187–199,200], are shown as grey circles within their geographical country or region of isolation, and the size of each circle is proportional to the number of strains sequenced from each location (number indicated when more than one genome is shown). The areas highlighted in brown are regions that contain active plague foci as determined by contemporary or historical data. yBP, years before present. Adapted with permission from the 'Global distribution of natural plague foci as of March 2016' from https://www.who.int/csr/disease/plague/Plague-map-2016.pdf.

foregut, the proventriculus (FIG. 5). This phenotype is determined by the unique acquisition and activity of certain genomic loci in *Y. pestis*, namely, the *Yersinia* murine toxin (*ymt*) gene, which is present on the pMT1 plasmid[140,141] and facilitates colonization of the arthropod midgut[141]. In addition, it is dependent on the pseudogenization of certain genes, namely, the biofilm downregulators *rcsA*, *PDE2* (also known as *rtn*), *PDE3* (also known as *y3389*)[141] and the urease gene *ureD*[148,149], which are, by contrast, active in *Y. pseudotuberculosis*. The biofilm prevents a blood meal from entering the flea's digestive tract, leaving it starving; as a result, the insect intensifies its feeding behaviour and promotes bacterial transmission to uninfected hosts[150–152]. This continuous transmission cycle among fleas and rodents, also called the enzootic phase of maintenance (FIG. 5), is thought to drive the preservation of plague foci around the world and is dependent on environmental and climatic factors as well as on host population densities[142,153–155]. Disruption of this equilibrium for reasons that are not well understood can cause disease eruption among susceptible rodent species, leading to so-called plague epizootics[142] (FIG. 5).

During that time, marked reductions in the rodent populations force fleas to seek alternative hosts, which can lead to infections in humans and, as a result, trigger the initiation of epidemics or pandemics.

Plague manifestation in humans has three disease forms, namely, bubonic, pneumonic and septicaemic[156]. Bubonic plague is the most common form of the disease and can cause up to 60% mortality when left untreated[157]. Subsequent to the bite of an infected flea, bacteria travel to the closest lymph node, where excessive replication occurs, giving rise to large swellings, the so-called buboes. In addition, following primary bubonic plague, bacteria can disseminate into the bloodstream to cause septicaemia (secondary septicaemic plague) and to the lungs, causing secondary pneumonic disease. Both forms are highly lethal disease presentations and cause nearly 100% mortality when left untreated. Only the pneumonic form can result in direct human-to-human transmission.

**Early evolution: plague in prehistory.** The time of divergence between *Y. pestis* and *Y. pseudotuberculosis* has been difficult to determine given the wide temporal interval produced by recent molecular dating attempts
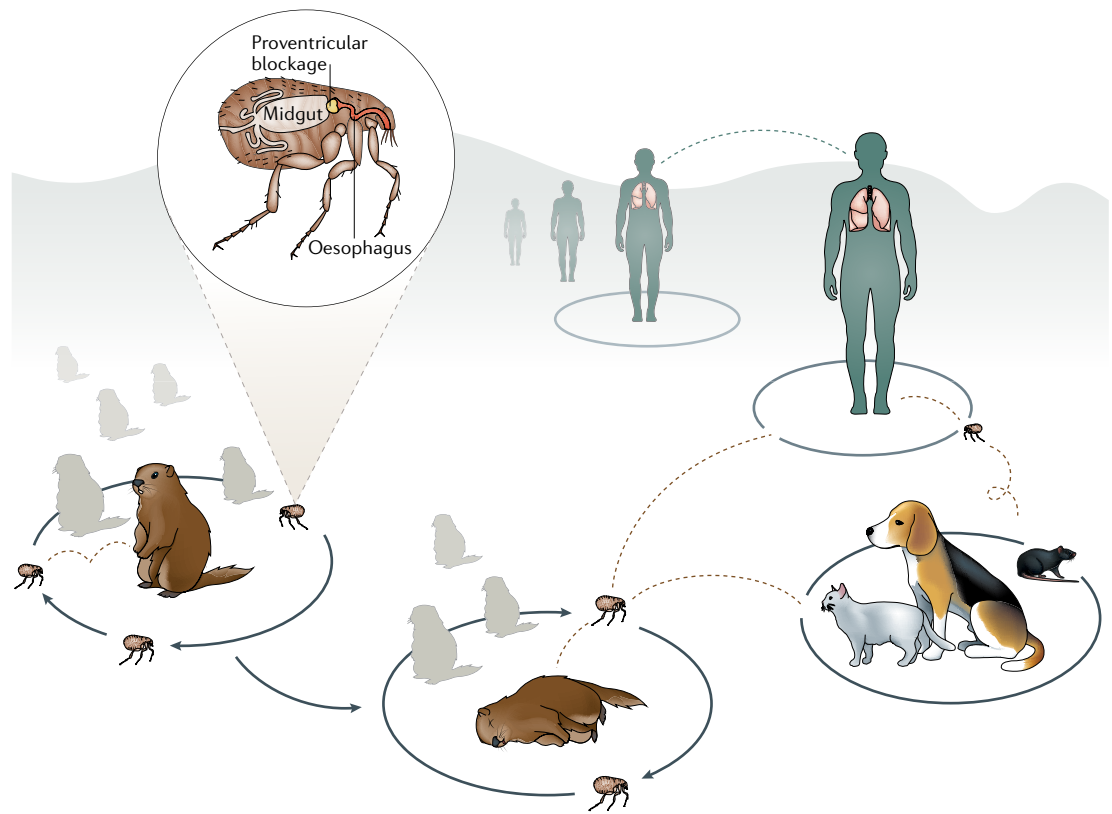
Fig. 5 | **Yersinia pestis ecology and transmission cycle.** A simplified version of the *Yersinia pestis* enzootic cycle, during which the bacterium is maintained among wild rodent populations through a flea-dependent transmission mechanism. Under poorly understood circumstances, plague epizootics, which are best explained as animal epidemics, can occur among susceptible rodent populations. During those periods, humans and other mammals are at highest risk of becoming infected with *Y. pestis*. Plague can manifest in humans in the bubonic, pneumonic and septicaemic forms. Pneumonic plague is the only form that can result in airborne transmission between humans.

based on aDNA data (13,000–79,000 years before present (yBP))[33,34]. Nevertheless, *Y. pestis* identification in human remains from Neolithic and Bronze Age Eurasia suggests that it caused human infections during these periods and originated more than 5,000 years ago[31,33,34]. These data have revealed important details about the early evolution of the bacterium. Genomic and phylogenetic analyses have shown that strains from the Late Neolithic and Bronze Age (LNBA) occupy a basal lineage in the *Y. pestis* phylogeny, and a recent study suggests the presence of even more basal variants in Neolithic Europe[31] (FIG. 6). Such analyses have demonstrated that, during its early evolution, the bacterium had not yet acquired important virulence factors consistent with the complex transmission cycle common to historical and extant strains. One of these genes is *ymt*, whose absence has been associated with an inability for flea midgut colonization in *Y. pestis*[141]. In addition, these strains possess the active forms of the *rcsA*, *PDE3*, *PDE2* and *ureD* genes, which suggests an impaired ability towards biofilm formation and blockage of the flea's proventriculus[141,149]. Finally, they possess an active flagellin gene (*flhD*), which is present as a pseudogene in all other *Y. pestis*, as it is a potent inducer of the innate immune response of the host[158]. As a result, during its initial evolutionary stages, *Y. pestis* may have been unable to efficiently transmit via a flea vector. Flea-borne transmission of *Y. pestis* is a known

prerequisite for bubonic plague development[141]; hence, it has been suggested that this disease phenotype was not present during prehistoric times[33,159]. In addition, these results have raised uncertainty regarding the possible vector and host mammalian species of the bacterium. The Bronze Age in Eurasia was a period of intense human migrations, which shaped the genomic landscape of modern-day Europe[85,160]. Remarkably, the *Y. pestis* LNBA lineage was shown to mirror human movements during that time[34] and was found in regions that do not host wild reservoir populations today (FIG. 4). The wide geographical distribution of these strains, their supposed limited bubonic disease potential and their relationship with human migration routes might together be indicative of a different reservoir host species compared to wild rodents that have a central role in plague transmission in areas such as Central and East Asia, where the disease is endemic today.

Nevertheless, an alternative mode of flea transmission, termed the early phase transmission, which occurs during the initial phases of infection and was suggested to be biofilm-independent[161], should also be considered as a possible way of *Y. pestis* propagation during its early evolution[34]. Although this transmission mechanism is currently not well understood, its comparative mode and efficiency in different rodent species have recently started to be assessed[162]. The oldest *Y. pestis* genomic
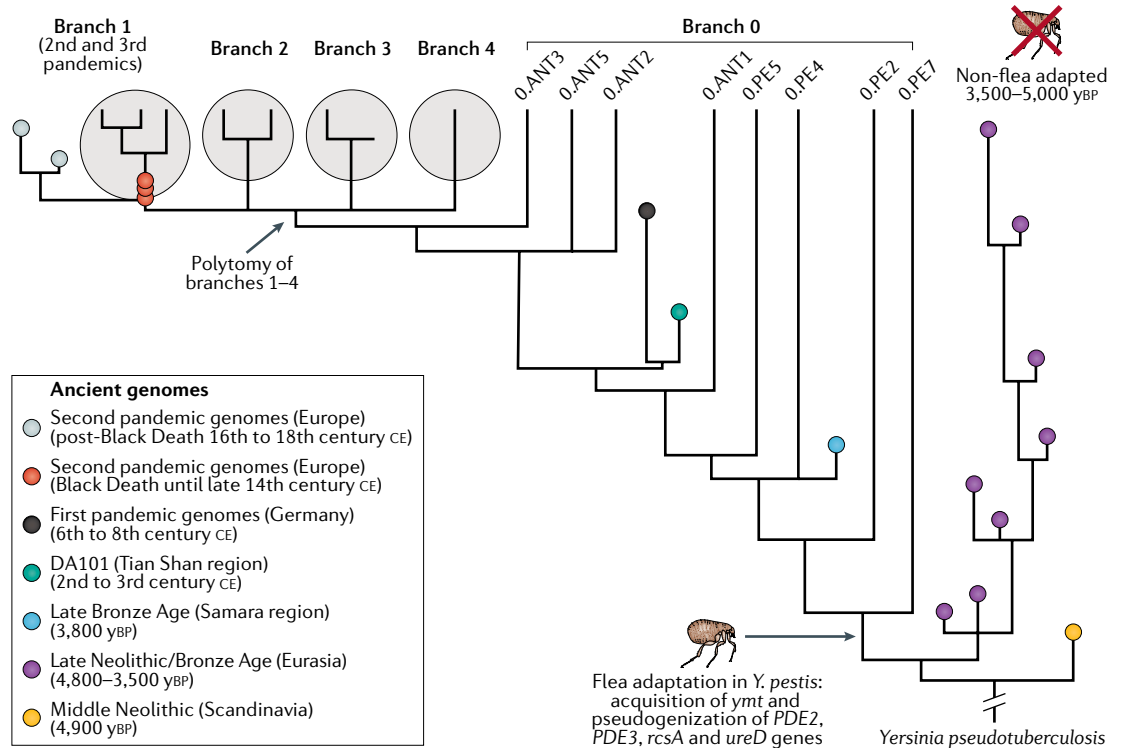
Fig. 6 | **Evolutionary history of *Yersinia pestis*.** A phylogenetic tree graphic depicting the evolutionary history of *Yersinia pestis* based on both ancient and modern genomes. Ancient strains that have been previously characterized by phylogenetic analysis are represented with coloured circles among the tree branches as follows: a Middle Neolithic genome is shown in yellow; Late Neolithic and Bronze Age (LNBA) genomes are shown in purple; a Late Bronze Age genome (RT5) encompassing signatures of flea adaptation is shown in blue; a pre-Justinian, 2nd century of the current era (CE), genome is shown in green; first-plague-pandemic genomes are shown in black; second plague pandemic, 14th-century genomes are shown in red; and post-Black Death (up until 18th century CE) genomes are shown in grey. Modern lineages are simplified and shown as branches of equal length in order to enhance the clarity of the graphic. The geographical distribution of modern strains is as follows (using universal country abbreviations): branch 1 (UGA, DRC, KEN, DZA, MDG, CHN, IND, IDN, MNM, USA and PER), branch 2 (RUS, AZE, KAZ, KGZ, UZB, TKM, CHN, IRN and NPL), branch 3 (CHN and MNG), branch 4 (RUS and MNG) and branch 0, including lineages 0.ANT3 (CHN and KGZ), 0.ANT5 (KGZ and KAZ), 0.ANT2 (CHN), 0.ANT1 (CHN), 0.PE5 (MNG), 0.PE4 (TJK, UZB, KGZ, RUS, CHN and MNG), 0.PE2 (GEO, ARM, AZE and RUS) and 0.PE7 (CHN). yBP, years before present.

evidence showing the full capacity for flea colonization similar to modern and historic strains was identified in two 3,800-year-old skeletons from the Samara region of modern-day Russia[37]. Although this strain was shown to occupy a phylogenetic position among modern *Y. pestis* lineages (FIG. 6), molecular dating analysis indicated that it originated ~4,000 years ago, suggesting that it overlapped temporally with the other Bronze Age strains that lacked the genetic prerequisites for arthropod transmission. Similar characteristics were previously identified in a low-coverage 3,000-year-old isolate from modern-day Armenia[33], which suggests that multiple forms of the bacterium were circulating in Eurasia between 5,000 and 3,000 years ago that may have had different transmission cycles and produced different disease phenotypes. As the propagation mechanisms of those strains are still uncertain, and the exact timing of flea-adaptation in *Y. pestis* is unknown, additional metagenomic screening from human and animal remains may provide relevant information on disease reservoirs and hosts across Neolithic and Bronze Age Eurasia.

It is becoming increasingly apparent that, aside from plague, other infectious diseases, such as those caused by HBV[43,44] and B19V[45] (TABLE 1), were circulating during the same time periods. Further pathogen screening coupled with a temporal assessment of human immune-associated genomic variants[84] may reveal key aspects of disease prevalence and susceptibility during this pivotal period of human history.

***Molecular insights from three historical plague pandemics.*** After the Bronze Age, bubonic plague has been associated with three historically recorded pandemics. The earliest accounts of the so-called first plague pandemic, which began with the Plague of Justinian (541 CE), suggest that it erupted in northern Africa in the mid-6th century CE[163,164] and subsequently spread through Europe and the vicinity until ~750 CE. The second historically recorded plague pandemic began with the infamous Black Death (1346–1353 CE)[96] and continued with outbreaks in Europe until the 18th century CE. The most recent third plague pandemic began in the mid-19th century in the Yunnan province of China, and it was during that time that Alexandre E. J. Yersin first described the bacterium in Hong Kong, in 1894 (FIG. 1). The third pandemic spread worldwide via marine routes

and has persisted until today in active foci in Africa, Asia and the Americas. Although the majority of modern plague cases derive from strains disseminated in this global dispersal, the pandemic is considered to have largely subsided since the 1950s[165].

The association of *Y. pestis* with the two earlier pandemics has, until recent years, been contentious. On the basis of their serological characterization, modern strains were traditionally grouped into three distinct biovars, namely, 'antiqua', 'medievalis' and 'orientalis', according to their ability to ferment glycerol and reduce nitrate[165,166]. In addition, historical accounts of the disease seemed to correlate with the supposed distinct geographical distributions of these biovars[166], and their phylogenetic relationships, as inferred from MLST data, reinforced the hypothesis that each was responsible for a single pandemic[136]. By contrast, later studies identified additional, atypical biovars[167], and more robust phylogenetic analysis suggested that phylogeography does not correlate clearly with the phenotypic distinctions described between these bacterial populations[92,130,168].

Recent genomic analyses have revealed high genetic diversity of the bacterium in East Asia, which invariably led to the assumption that *Y. pestis* emerged there[130]. However, a strong research focus on the diversity of the bacterium in these endemic regions, mainly China, has contributed to a profound sampling bias in the available modern data (FIG. 4). More recent investigations have revealed previously uncharacterized genetic diversity in the Caucasus region and in the central Asian steppe that ought to be further explored[169–172] (FIG. 4). Currently, the evolutionary tree of the bacterium is characterized by five main phylogenetic branches (FIG. 6). The most ancestral, branch 0, includes strains distributed across China, Mongolia and the areas encompassing the former Soviet Union. The more phylogenetically derived branches 1–4 were formed through a rapid population expansion event and are today found in Asia, Africa and the Americas[130]. Their wide distribution mainly reflects the geographical breadth of branch 1, which is associated with the third plague pandemic that spread worldwide during the 19th and 20th centuries[92] and is still responsible for more confined epidemics such as those reported in Madagascar[173].

The analysis of aDNA from historical epidemic contexts has generated important information regarding the evolutionary history of plague. The recovery of *Y. pestis* DNA via PCR from remnants of human dental pulp suggested the involvement of the bacterium in both the first and second pandemics; however, these results were difficult to authenticate[8,174,175]. Subsequent PCR-based SNP typing of ancient specimens offered some phylogenetic resolution and revealed an expected ancestral placement of medieval strains in the *Y. pestis* phylogeny[62–64]. More recently, full characterization and authentication of the bacterium were achieved using plasmid and whole-genome enrichment coupled with NGS[17,20,35,36].

Historical accounts of the first plague pandemic (6th to 8th centuries CE) suggest that the disease expanded mainly across the Mediterranean basin; however, its exact breadth and impact have been difficult to assess given the limited availability of historical and archaeological data, with the latter being currently under revision[176]. Two recent studies have reconstructed 6th-century *Y. pestis* genomes from southern Germany[35,36] (FIG. 4), a region that lacked historical documentation of the pandemic. Phylogenetic analysis showed that both genomes belong to a lineage that is today extinct and is closely related to strains from modern-day China[35,36], which suggests the possibility of an East Asian origin of the first pandemic. This hypothesis was recently reinforced by the publication of a 2nd-century to 3rd-century *Y. pestis* genome from the Tian Shan mountains of modern-day Kyrgyzstan[39], which shares a common ancestor with the Justinianic-plague lineage (FIGS 4,6). However, given the >300-year age difference between these strains[35,36,39], as well as the aforementioned East Asian sampling bias of modern *Y. pestis* data[130], the geographical origin of the pandemic remains hypothetical. Retrieval of additional *Y. pestis* strain diversity from that time period, particularly from areas known to have played an important role in the entry of this bacterium into Europe, that is, the eastern Mediterranean region, may hold clues about its putative source.

The beginning of the second plague pandemic, 600 years later, was marked by the notorious Black Death of Europe (1346–1353 CE), estimated to have caused an up to 60% reduction of the continental population in only 5 years[96]. Historical records suggest that the first outbreaks occurred in the Lower Volga region of Russia, and the disease then spread into southern Europe through the Crimean peninsula[96]. Initial analysis of *Y. pestis* via PCR from victims of the Black Death revealed a distinct phylogenetic positioning of two mid-to-late-14th-century strains and led to the proposal that the disease entered the continent through independent pulses[64]. By contrast, whole-genome analysis of ancient strains from western, northern and southern Europe demonstrated a lack of *Y. pestis* diversity during the Black Death, which suggests its fast spread through the continent and favours a single-wave entry model of the bacterium into Europe[20,30,38], although the possible presence of additional strain diversity during that time has recently been explored[30]. Intriguingly, the phylogenetic positioning of the Black Death *Y. pestis* genomes places them on branch 1, only two nucleotide substitutions away from the 'star-like' diversification of branches 1–4 (FIG. 6), which gave rise to most of the strain diversity identified around the world today[38,130].

After the Black Death, plague epidemics continued to affect Europe until the 18th century[177,178]. Inferred climatic data from tree ring records in central Asia and Europe have recently suggested that such epidemics were likely caused by multiple introductions of the bacterium into Europe as a result of climate-driven disruptions of pre-existing Asian reservoirs[179]. By contrast, ancient genetic and genomic evidence supports the persistence of the disease in Europe for 400 years after the Black Death[32,38,62]. Analysis of *Y. pestis* strains spanning from the late 14th to the 18th century CE has revealed the formation of at least two European lineages that were responsible for the ensuing medieval epidemics (FIG. 6). Both lineages derive from the Black

Death *Y. pestis* strain identified in 14th-century western, northern and southern Europe[30,32,38], suggesting that they likely arose locally. The first lineage survives today and gave rise to modern branch 1 strains[30,38] (which are associated with the third plague pandemic), suggesting the European Black Death as a source for modern-day epidemics[38]. The second lineage has not been identified among present-day diversity and currently encompasses strains from 16th-century Germany[38] and 18th-century France (Great Plague of Marseille, 1720–1722 CE) (FIG. 6). These phylogenetic patterns are consistent with a continuous persistence of the bacterium in Europe during the second plague pandemic. In addition, they are supported by analyses of historical records that suggest the existence of plague reservoirs in the continent until the 18th century CE[180].

*Y. pestis* is absent from most of Europe today; specifically, no active foci exist west of the Black Sea. Plague is thought to have disappeared from most of Europe at the end of the second pandemic (18th century CE). This finding is striking given the thousands of outbreaks that were recorded in the continent until that time[177,178]. The reasons for its disappearance are unknown, although numerous hypotheses have been put forward[181], including a change in domestic rodent populations in Europe, namely, the replacement of the black rat, *Rattus rattus*, by the brown rat, *Rattus norvegicus*[181]; an acquired plague immunity among humans and/or rodents[181] (although this hypothesis requires an update to accommodate the recent identification of *Y. pestis* in Europe 5,000 years ago[31,33,34] and the involvement of the bacterium in the first plague pandemic[35,36]); the increased living standards such as the better nutrition and hygienic conditions at the beginning of the Early Modern Era, which may have contributed to improved overall health conditions in Europe and likely decreased the number of rats and ectoparasites in human environments[181,182]; and the potential disruption of the European wild rodent ecological niche owing to habitat loss and industrialization starting in 1700 CE[180]. Given the contribution that molecular data can offer in these discussions, future research on ancient sources of *Y. pestis* DNA will be instrumental in further revealing the history of one of humankind's most devastating pathogens.

## Conclusions

The analysis of ancient pathogen genomes has afforded promising views into past infectious disease history. For *Y. pestis*, aDNA exploration of its evolutionary past has revealed how a predominantly environmental bacterium and opportunistic gastroenteric pathogen developed into an extremely virulent form by acquisition of only a few virulence factors. We eagerly await revelations on a similar scale for other important pathogens that are expected to arise from deep temporal sampling and genomic reconstruction, as made possible through the recent advancements discussed here.

Integration of ancient pathogen genomes into disease modelling and human population genetic frameworks, as well as their analysis alongside the information offered by the archaeological, historical and palaeopathological records, will help build a more interdisciplinary and complete picture of host–pathogen interactions and human evolutionary history over time.

Published online 5 April 2019

1. Armelagos, G. J., Goodman, A. H. & Jacobs, K. H. The origins of agriculture: population growth during a period of declining health. *Popul. Environ.* **13**, 9–22 (1991).
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. Ortner, D. J. *Identification of Pathological Conditions in Human Skeletal Remains* 2nd edn (Academic Press, 2003).
4. Buikstra, J. E. & Roberts, C. *The Global History of Paleopathology: Pioneers and Prospects* (Oxford Univ. Press, 2012).
5. Arriaza, B. T., Salo, W., Aufderheide, A. C. & Holcomb, T. A. Pre-Columbian tuberculosis in Northern Chile: molecular and skeletal evidence. *Am. J. Phys. Anthropol.* **98**, 37–45 (1995).
6. Salo, W. L., Aufderheide, A. C., Buikstra, J. & Holcomb, T. A. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *Proc. Natl Acad. Sci. USA* **91**, 2091–2094 (1994).
7. Zink, A., Haas, C. J., Reischl, U., Szeimies, U. & Nerlich, A. G. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J. Med. Microbiol.* **50**, 355–366 (2001).
8. Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O. & Raoult, D. Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. *Proc. Natl Acad. Sci. USA* **95**, 12637–12640 (1998).
9. Spigelman, M. & Lemma, E. The use of the polymerase chain reaction (PCR) to detect *Mycobacterium tuberculosis* in ancient skeletons. *Int. J. Osteoarchaeol.* **3**, 137–143 (1993).
10. Pääbo, S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl Acad. Sci. USA* **86**, 1939–1943 (1989).

11. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* **7**, e34131 (2012).
12. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
    **The study provides a quantitative description of aDNA-associated patterns of nucleotide misincorporation and fragmentation that are currently used as primary authentication criteria.**
13. Cooper, A. & Poinar, H. N. Ancient DNA: do it right or not at all. *Science* **289**, 1139–1139 (2000).
14. Gilbert, M. T. P. et al. Absence of *Yersinia pestis*-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology* **150**, 341–354 (2004).
15. Shapiro, B., Rambaut, A. & Gilbert, M. T. P. No proof that typhoid caused the Plague of Athens (a reply to Papagrigorakis et al.). *Int. J. Infect. Dis.* **10**, 334–335 (2006).
16. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376 (2005).
17. Schuenemann, V. J. et al. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* **108**, E746–E752 (2011).
18. Green, R. E. et al. The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**, 2494–2502 (2009).
19. Key, F. M., Posth, C., Krause, J., Herbig, A. & Bos, K. I. Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* **33**, 508–520 (2017).
20. Bos, K. I. et al. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
    **The study describes the first whole-genome sequence of an ancient bacterial pathogen through the use of high-throughput sequencing.**

21. Pääbo, S. et al. Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645–679 (2004).
22. Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567 (2013).
23. Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLOS ONE* **10**, e0129102 (2015).
24. Hansen, H. B. et al. Comparing ancient DNA preservation in petrous bone and tooth cementum. *PLOS ONE* **12**, e0170940 (2017).
25. Margaryan, A. et al. Ancient pathogen DNA in human teeth and petrous bones. *Ecol. Evol.* **8**, 3534–3542 (2018).
26. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
27. Schuenemann, V. J. et al. Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. *PLOS Pathog.* **14**, e1006997 (2018).
28. Schuenemann, V. J. et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* **341**, 179–183 (2013).
    **The study presents the first de novo assembled ancient pathogen genome and an analysis of *M. leprae* in medieval Europe.**
29. Schuenemann, V. J. et al. Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains. *PLOS Negl. Trop. Dis.* **12**, e0006447 (2018).
30. Namouchi, A. et al. Integrative approach using Yersinia pestis genomes to revisit the historical landscape of plague during the Medieval Period. *Proc. Natl Acad. Sci. USA* **115**, E11790–E11797 (2018).
31. Rascovan, N. et al. Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic Decline. *Cell* **176**, 295–305 (2018).
32. Bos, K. I. et al. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, e12994 (2016).

33. Rasmussen, S. et al. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
**The study describes *Y. pestis* genomes from Bronze Age human remains and provides a chronological timing of virulence determinant acquisition during the early evolution of the bacterium.**

34. Andrades Valtueña, A. A. et al. The Stone Age plague and its persistence in Eurasia. *Curr. Biol.* **27**, 3683–3691 (2017).

35. Feldman, M. et al. A high-coverage *Yersinia pestis* genome from a sixth-century justinianic plague victim. *Mol. Biol. Evol.* **33**, 2911–2923 (2016).

36. Wagner, D. M. et al. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
**The study reports the first genome-wide analysis of *Y. pestis* from victims of the Plague of Justinian, directly implicating the bacterium in the first plague pandemic.**

37. 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).
**This study presents the oldest genomic evidence of flea adaptation in *Y. pestis.***

38. Spyrou, M. A. et al. Historical *Y. pestis* genomes reveal the European Black Death as the source of ancient and modern plague pandemics. *Cell Host Microbe* **19**, 874–881 (2016).

39. de Barros Damgaard, P. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369 (2018).

40. Guellil, M. et al. Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia. *Proc. Natl Acad. Sci. USA* **115**, 10422–10427 (2018).

41. Vågene, A. J. et al. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
**This paper presents the metagenomic tool MALT and is the first case study to demonstrate metagenomic detection of ancient pathogens in the absence of prior knowledge on the causative agent of an epidemic.**

42. Marciniak, S. et al. *Plasmodium falciparum* malaria in 1st–2nd century CE southern Italy. *Curr. Biol.* **26**, R1220–R1222 (2016).

43. Mühlemann, B. et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).

44. Krause-Kyora, B. et al. Neolithic and Medieval virus genomes reveal complex evolution of hepatitis B. *eLife* **7**, e36666 (2018).
**The studies by Mühlemann (Nature, 2018) and Krause-Kyora (eLife, 2018) present a time transect of HBV genomes, spanning from the Neolithic period to the medieval period, and provide an overview of the HBV population history across millennia.**

45. Mühlemann, B. et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proc. Natl Acad. Sci. USA* **115**, 7557–7562 (2018).

46. Warinner, C. et al. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336 (2014).
**This study provides an analysis of the composition of human dental calculus from ancient individuals, showing the presence of oral microbiome bacterial DNA, periodontal pathogen DNA and proteins associated with host immunity.**

47. Kay, G. L. et al. Recovery of a medieval *Brucella melitensis* genome using shotgun metagenomics. *mBio* **5**, e01337–14 (2014).

48. Devault, A. M. et al. A molecular portrait of maternal sepsis from Byzantine Troy. *eLife* **6**, e20983 (2017).

49. Maixner, F. et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* **351**, 162–165 (2016).
**The study provides insights into the genomic history of *H. pylori* over several millennia through a population genomic analysis of a Copper Age strain against a worldwide data set.**

50. Duggan, A. T. et al. 17th century variola virus reveals the recent history of smallpox. *Curr. Biol.* **26**, 3407–3412 (2016).

51. Biagini, P. et al. Variola virus in a 300-year-old Siberian mummy. *N. Engl. J. Med.* **367**, 2057–2059 (2012).

52. Kay, G. L. et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat. Commun.* **6**, 6717 (2015).

53. Ross, Z. P. et al. The paradox of HBV evolution as revealed from a 16th century mummy. *PLOS Pathog.* **14**, e1006750 (2018).

54. Kahila Bar-Gal, G. et al. Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* **56**, 1671–1680 (2012).

55. Devault, A. M. et al. Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N. Engl. J. Med.* **370**, 334–340 (2014).

56. Gelabert, P. et al. Mitochondrial DNA from the eradicated European *Plasmodium vivax* and *P. falciparum* from 70-year-old slides from the Ebro Delta in Spain. *Proc. Natl Acad. Sci. USA* **113**, 11495–11500 (2016).

57. Worobey, M. et al. 1970s and 'patient 0'HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* **539**, 98 (2016).

58. Taubenberger, J. K. et al. Characterization of the 1918 influenza virus polymerase genes. *Nature* **437**, 889 (2005).

59. Yoshida, K. et al. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* **2**, e00731 (2013).

60. Martin, M. D. et al. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nat. Commun.* **4**, 2172 (2013).

61. Harkins, K. M. et al. Screening ancient tuberculosis with qPCR: challenges and opportunities. *Phil. Trans. R. Soc. B* **370**, 20130622 (2015).

62. Seifert, L. et al. Genotyping *Yersinia pestis* in historical plague: evidence for long-term persistence of *Y. pestis* in Europe from the 14th to the 17th century. *PLOS ONE* **11**, e0145194 (2016).

63. Harbeck, M. et al. *Yersinia pestis* DNA from skeletal remains from the 6th century AD reveals insights into Justinianic Plague. *PLOS Pathog.* **9**, e1003349 (2013).

64. Haensch, S. et al. Distinct clones of *Yersinia pestis* caused the black death. *PLOS Pathog.* **6**, e1001134 (2010).

65. Bos, K. I. et al. Parallel detection of ancient pathogens via array-based DNA capture. *Phil. Trans. R. Soc. B* **370**, 20130375 (2015).

66. Devault, A. M. et al. Ancient pathogen DNA in archaeological samples detected with a microbial detection array. *Sci. Rep.* **4**, 4245 (2014).

67. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

68. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

69. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2015).

70. Segata, N. et al. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811 (2012).

71. Louvel, G., Der Sarkissian, C., Hanghøj, K. & Orlando, L. metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data. *Mol. Ecol. Resour.* **16**, 1415–1427 (2016).

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

73. Warinner, C. et al. A robust framework for microbial archaeology. *Annu. Rev. Genom. Hum. Genet.* **18**, 321–356 (2017).

74. Müller, R., Roberts, C. A. & Brown, T. A. Complications in the study of ancient tuberculosis: Presence of environmental bacteria in human archaeological remains. *J. Archaeol. Sci.* **68**, 5–11 (2016).

75. Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. v. & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799 (2001).

76. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

77. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).

78. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B* **370**, 20130624 (2015).

79. Hodges, E. et al. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat. Protoc.* **4**, 960–974 (2009).

80. Burbano, H. A. et al. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* **328**, 723–725 (2010).

81. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).

82. Ávila-Arcos, M. C. et al. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Sci. Rep.* **1**, 74 (2011).

83. Cruz-Dávalos, D. I. et al. Experimental conditions improving in-solution target enrichment for ancient DNA. *Mol. Ecol. Resour.* **17**, 508–522 (2017).

84. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).

85. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
**This study delineates large-scale population migrations into Europe during the Bronze Age by analysis of human genome-wide data of individuals living between 8,000 and 3,000 �𝘣ᴘ.**

86. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).

87. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).

88. Posth, C. et al. Language continuity despite population replacement in Remote Oceania. *Nat. Ecol. Evol.* **2**, 731 (2018).

89. Rasmussen, M. et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).

90. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).

91. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).

92. Morelli, G. et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).

93. Comas, I. et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* **45**, 1176 (2013).
**The study shows a possible co-expansion of *M. tuberculosis* among human populations during out-of-Africa migrations.**

94. Roberts, C. A. & Buikstra, J. E. *The Bioarchaeology of Tuberculosis: a Global Perspective on a Re-Emerging Disease* (Univ. Press of Florida, 2003).

95. Cohn, S. K. Jr. *The Black Death Transformed: Disease and Culture in Early Renaissance Europe* (Arnold, 2002).

96. Benedictow, O. J. *The Black Death, 1346-1353: The Complete History* (Boydell & Brewer, 2004).

97. Ortner, D. J. in *Advances in Human Palaeopathology* (eds Pinhasi, R. & Mays, S.) 189–214 (John Wiley & Sons, 2008).

98. Cunha, C. B. & Cunha, B. A. in *Paleomicrobiology: Past Human Infections* (eds Raoult, D. & Drancourt, M.) 1–20 (Springer, 2008).

99. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

100. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).

101. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

102. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).

103. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

104. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2005).

105. Krause-Kyora, B. et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nat. Commun.* **9**, 1569 (2018).

106. Robbins, G. et al. Ancient skeletal evidence for leprosy in India (2000 BC). *PLOS ONE* **4**, e5669 (2009).

107. Köhler, K. et al. Possible cases of leprosy from the Late Copper Age (3780–3650 cal BC) in Hungary. *PLOS ONE* **12**, e0185966 (2017).

108. Wong, S. H. et al. Leprosy and the adaptation of human toll-like receptor 1. *PLOS Pathog.* **6**, e1000979 (2010).

109. Zhou, Z. et al. Pan-genome analysis of ancient and modern *Salmonella enterica* demonstrates genomic stability of the invasive para C lineage for millennia. *Curr. Biol.* **28**, 2420–2428 (2018).
   **The study describes a 12th century *Salmonella enterica* subsp. *enterica* serovar Paratyphi C genome and its analysis alongside a comprehensive data set of thousands of *S. enterica* strains.**

110. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLOS Comp. Biol.* **11**, e1004041 (2015).

111. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).

112. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).

113. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLOS Genet.* **8**, e1002453 (2012).

114. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).

115. Monot, M. et al. On the origin of leprosy. *Science* **308**, 1040–1042 (2005).

116. Wetterstrand, K. A. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). *genome.gov* https://www.genome.gov/sequencingcostsdata/ (updated 25 Apr 2018).

117. Wattam, A. R. et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581–D591 (2013).

118. Alikhan, N.-F., Zhou, Z., Sergeant, M. J. & Achtman, M. A genomic overview of the population structure of *Salmonella. PLOS Genet.* **14**, e1007261 (2018).
   **The paper introduces a web-based platform that performs genome assembly and multilocus sequence typing analysis and can be used for the retrieval of large data sets on enteric bacteria.**

119. Zhou, Z. et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* **28**, 1395–1404 (2018).

120. Duchêne, S. et al. Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* **2**, e000094 (2016).

121. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).

122. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).

123. Chan, J. Z.-M. et al. Metagenomic analysis of tuberculosis in a mummy. *N. Engl. J. Med.* **369**, 289–290 (2013).
   **The study describes the first sequenced ancient *M. tuberculosis* genome and shows the presence of mixed infections in 18th century Europe.**

124. Duchêne, S., Duchêne, D., Holmes, E. C. & Ho, S. Y. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol. Biol. Evol.* **32**, 1895–1906 (2015).

125. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

126. Ferreira, M. A. & Suchard, M. A. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can. J. Stat.* **36**, 355–368 (2008).

127. Maturana Russel, P., Brewer, B. J., Klaere, S. & Bouckaert, R. R. Model selection and parameter inference in phylogenetics using nested sampling. *Syst. Biol.* **68**, 219–233 (2019).

128. Baele, G. et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).

129. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLOS Biol.* **4**, e88 (2006).

130. Cui, Y. et al. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis. Proc. Natl Acad. Sci. USA* **110**, 577–582 (2013).
   **The study presents a comprehensive *Y. pestis* modern genomic data set from East Asia and demonstrates extensive clock-rate variations across the *Y. pestis* phylogeny.**

131. Kingman, J. F. C. The coalescent. *Stoch. Process Their Appl.* **13**, 235–248 (1982).

132. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).

133. Stadler, T., Kuhnert, D., Bonhoeffer, S. & Drummond, A. J. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl Acad. Sci. USA* **110**, 228–233 (2013).

134. Stadler, T. et al. Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol.* **29**, 347–357 (2011).

135. To, T.-H., Jung, M., Lycett, S. & Gascuel, O. Fast dating using least-squares criteria and algorithms. *Syst. Biol.* **65**, 82–97 (2015).

136. Achtman, M. et al. *Yersinia pestis*, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. *Proc. Natl Acad. Sci. USA* **96**, 14043–14048 (1999).
   **This study is the first to show that *Y. pestis* is a recently emerged clone of *Y. pseudotuberculosis.***

137. Chain, P. S. et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with Yersinia pseudotuberculosis. *Proc. Natl Acad. Sci. USA* **101**, 13826–13831 (2004).

138. Zimbler, D. L., Schroeder, J. A., Eddy, J. L. & Lathem, W. W. Early emergence of *Yersinia pestis* as a severe respiratory pathogen. *Nat. Commun.* **6**, 7487 (2015).

139. Hinnebusch, J. et al. Murine toxin of *Yersinia pestis* shows phospholipase D activity but is not required for virulence in mice. *Int. J. Med. Microbiol.* **290**, 483–487 (2000).

140. Hinnebusch, B. J. et al. Role of Yersinia murine toxin in survival of *Yersinia pestis* in the midgut of the flea vector. *Science* **296**, 733–735 (2002).

141. Sun, Y. C., Jarrett, C. O., Bosio, C. F. & Hinnebusch, B. J. Retracing the evolutionary path that led to flea-borne transmission of *Yersinia pestis. Cell Host Microbe* **15**, 578–586 (2014).
   **This study presents a functional description of genes associated with flea-dependent colonization and transmission in *Y. pestis.***

142. Gage, K. L. & Kosoy, M. Y. in *Recovery of the Black-Footed Ferret Progress and Continuing Challenges. United States Geological Survey Scientific Investigations Report 5293* (eds Roelle, J. E., Miller, B. J., Godbey, J. L. & Biggins, D. E.) 213–231 (U.S. Geological Survey, 2006).

143. Gage, K. L. & Kosoy, M. Y. Natural history of plague: perspectives from more than a century of research. *Annu. Rev. Entomol.* **50**, 505–528 (2005).

144. Anisimov, A. P., Lindler, L. E. & Pier, G. B. Intraspecific diversity of *Yersinia pestis. Clin. Microbiol. Rev.* **17**, 434–464 (2004).

145. Riehm, J. M. et al. *Yersinia pestis* in small rodents, Mongolia. *Emerg. Infect. Dis.* **17**, 1320 (2011).

146. Hinnebusch, B. J., Bland, D. M., Bosio, C. F. & Jarrett, C. O. Comparative ability of oropsylla montana and xenopsylla cheopis fleas to transmit Yersinia pestis by two different mechanisms. *PLOS Negl. Trop. Dis.* **11**, e0005276 (2017).

147. Dean, K. R. et al. Human ectoparasites and the spread of plague in Europe during the second pandemic. *Proc. Natl Acad. Sci. USA* **115**, 1304–1309 (2018).

148. Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E. & Simonet, M. Silencing and reactivation of urease in *Yersinia pestis* is determined by one G residue at a specific position in the ureD gene. *Infect. Immun.* **69**, 170–176 (2001).

149. Chouikha, I. & Hinnebusch, B. J. Silencing urease: a key evolutionary step that facilitated the adaptation of *Yersinia pestis* to the flea-borne transmission route. *Proc. Natl Acad. Sci. USA* **111**, 18709–18714 (2014).

150. Hinnebusch, B. J. in *Advances in Yersinia Research* (eds de Almeida, A. M. P. & Leal, N. C.) Vol. 954 237–243 (Springer, 2012).

151. Jarrett, C. O. et al. Transmission of *Yersinia pestis* from an infectious biofilm in the flea vector. *J. Infect. Dis.* **190**, 782–792 (2004).

152. Bacot, A. W. & Martin, C. J. LXVII. Observations on the mechanism of the transmission of plague by fleas. *J. Hyg.* **13**, 423 (1914).

153. Keeling, M. J. & Gilligan, C. A. Metapopulation dynamics of bubonic plague. *Nature* **407**, 903–906 (2000).

154. Stenseth, N. C. et al. Plague dynamics are driven by climate variation. *Proc. Natl Acad. Sci. USA* **103**, 13110–13115 (2006).

155. Davis, S. et al. Predictive thresholds for plague in Kazakhstan. *Science* **304**, 736–738 (2004).

156. Perry, R. D. & Fetherston, J. D. *Yersinia pestis* — etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).

157. World Health Organization. Plague. *WHO* http://www.who.int/csr/disease/plague/en/ (2017).

158. Minnich, S. A. & Rohde, H. N. in *The Genus Yersinia: From Genomics to Function* (eds Perry, R. D. & Fetherston, J. D.) Vol. 603 298–311 (Springer, 2007).

159. Hinnebusch, B. J., Chouika, I. & Sun, Y. C. Ecological opportunity, evolution, and the emergence of flea-borne plague. *Infect. Immun.* **84**, 1932–1940 (2016).

160. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
   **This study presents human genome-wide data from the Bronze Age period that was used to delineate large-scale migrations across Eurasia. The same data set was later used for pathogen screening and could show the presence of *Y. pestis*, HBV and B19V in those populations.**

161. Eisen, R. J., Dennis, D. T. & Gage, K. L. The role of early-phase transmission in the spread of *Yersinia pestis. J. Med. Entomol.* **52**, 1183–1192 (2015).

162. Bland, D. M., Jarrett, C. O., Bosio, C. F. & Hinnebusch, B. J. Infectious blood source alters early foregut infection and regurgitative transmission of *Yersinia pestis* by rodent fleas. *PLOS Pathog.* **14**, e1006859 (2018).

163. Harper, K. *The Fate of Rome: Climate, Disease, and the End of an Empire* (Princeton Univ. Press, 2017).

164. Sarris, P. The Justinianic plague: origins and effects. *Contin. Chang.* **17**, 169–182 (2002).

165. Pollitzer, R. *Plague (World Health Organization Monograph Series; no. 22)* (World Health Organization, 1954).

166. Devignat, R. Varietes de l'espece Pasteurella pestis: nouvelle hypothese [French]. *Bull. World Health Organ.* **4**, 247–263 (1951).

167. Zhou, D. et al. Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, microtus. *J. Bacteriol.* **186**, 5147–5152 (2004).

168. Achtman, M. et al. Microevolution and history of the plague bacillus, *Yersinia pestis. Proc. Natl Acad. Sci. USA* **101**, 17837–17842 (2004).

169. Kutyrev, V. V. et al. Phylogeny and classification of Yersinia pestis through the lens of strains from the plague foci of commonwealth of independent states. *Front. Microbiol.* **9**, 1106 (2018).

170. Eroshenko, G. A. et al. *Yersinia pestis* strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. *PLOS ONE* **12**, e0187230 (2017).

171. Kislichkina, A. A. et al. Nineteen whole-genome assemblies of *Yersinia pestis* subsp. microtus, including representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica. *Genome Announc.* **3**, e01342–15 (2015).

172. Zhgenti, E. et al. Genome assemblies for 11 *Yersinia pestis* strains isolated in the caucasus region. *Genome Announc.* **3**, e01030–15 (2015).

173. Vogler, A. J. et al. Temporal phylogeography of *Yersinia pestis* in Madagascar: insights into the long-term maintenance of plague. *PLOS Negl. Trop. Dis.* **11**, e0005887 (2017).

174. Drancourt, M. et al. *Yersinia pestis* Orientalis in remains of ancient plague patients. *Emerg. Infect. Dis.* **13**, 332 (2007).

175. Drancourt, M. et al. Genotyping, Orientalis-like *Yersinia pestis*, and plague pandemics. *Emerg. Infect. Dis.* **10**, 1585 (2004).

176. McCormick, M. Tracking mass death during the fall of rome's empire (I). *J. Roman Archaeol.* **28**, 325–357 (2015).

177. Büntgen, U., Ginzler, C., Esper, J., Tegel, W. & McMichael, A. J. Digitizing historical plague. *Clin. Infect. Dis.* **55**, 1586–1588 (2012).

178. Biraben, J.-N. *Les Hommes et la Peste en France et dans les Pays Européens et Méditerranéens. Tome II. – Les Hommes Face à la Peste* (Mouton De Gruyter, 1976).

179. Schmid, B. V. et al. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc. Natl Acad. Sci. USA* **112**, 3020–3025 (2015).

180. Carmichael, A. G. in *Pandemic Disease in the Medieval World: Rethinking the Black Death* Vol. 1 (ed. Green, M. H.) 157–192 (ARC Medieval Press, 2014).

181. Appleby, A. B. The disappearance of plague: a continuing puzzle. *Econ. Hist. Rev.* **33**, 161–173 (1980).

182. DeWitte, S. N. Mortality risk and survival in the aftermath of the medieval Black Death. *PLOS ONE* **9**, e96513 (2014).

183. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).

184. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5448 (2010).

185. Gansauge, M.-T. et al. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017).

186. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

187. Kislichkina, A. A. et al. Six whole-genome assemblies of *Yersinia pestis* subsp. microtus bv. ulegeica (Phylogroup 0. PE5) strains isolated from mongolian natural plague foci. *Genome Announc.* **6**, e00536–18 (2018).

188. Song, Y. et al. Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res.* **11**, 179–197 (2004).

189. Cáceres, O. et al. Whole-genome sequencing and comparative analysis of *Yersinia pestis*, the causative agent of a plague outbreak in northern Peru. *Genome Announc.* **1**, e00249–12 (2013).

190. Dai, R. et al. Human plague associated with Tibetan sheep originates in marmots. *PLOS Negl. Trop. Dis.* **12**, e0006635 (2018).

191. Johnson, S. L. et al. Thirty-two complete genome assemblies of nine Yersinia species, including *Y. pestis*, Y. pseudotuberculosis, and Y. enterocolitica. *Genome Announc.* **3**, e00148–15 (2015).

192. Kislichkina, A. A. et al. Nine whole-genome assemblies of *Yersinia pestis* subsp. microtus bv. Altaica strains isolated from the altai mountain natural plague focus (No. 36) in Russia. *Genome Announc.* **6**, e01440–17 (2018).

193. Garcia, E. et al. in *The Genus Yersinia: From Genomics to Function* (eds Perry, R. D. & Fetherston, J. D.) Vol. 603 17–22 (Springer, 2007).

194. Chain, P. S. et al. Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J. Bacteriol.* **188**, 4453–4463 (2006).

195. Parkhill, J. et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523 (2001).

196. Mahale, K. N. et al. Draft genome sequences of *Yersinia pestis* strains from the 1994 plague epidemic of Surat and 2002 Shimla outbreak in India. *Indian J. Microbiol.* **54**, 480–482 (2014).

197. Gibbons, H. S. et al. Comparative genomics of 2009 seasonal plague (*Yersinia pestis*) in New Mexico. *PLOS ONE* **7**, e31604 (2012).

198. Shen, X. et al. Complete genome sequences of *Yersinia pestis* from natural foci in China. *J. Bacteriol.* **192**, 3551–3552 (2010).

199. Touchman, J. W. et al. A North American *Yersinia pestis* draft genome sequence: SNPs and phylogenetic analysis. *PLOS ONE* **2**, e220 (2007).

200. Malek, M. A. et al. *Yersinia pestis* halotolerance illuminates plague reservoirs. *Sci. Rep.* **7**, 40022 (2017).

201. Mendum, T. A. et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. *BMC Genomics* **15**, 270 (2014).

202. Tumpey, T. M. et al. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).