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.
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. 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 and the cementum layer of teeth 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.

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; *Mycobacterium leprae*, which could be isolated from portions of the maxilla and various long bones; and *Treponema pallidum* subsp. pallidum and *T. pallidum* subsp. pertenue, which have been isolated from long bones.

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 bloodstream infections do not typically produce diagnostic bone changes as opposed to those that affect their hosts chronically. Therefore, if infections have caused mortality in the acute phase, as is the case for individuals from epidemics and pandemics 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. 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*, *Borreliac recurrentis* and *Salmonellla enterica*; ancient eukaryotic pathogens such as *Plasmodium falciparum*; and ancient viruses such as hepatitis B virus (HBV) and human parvovirus B19 (B19V). Even *M. leprae*, which commonly manifests in the chronic form, has been retrieved from ancient teeth.

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*; calcified nodules, which have yielded whole genomes from *Brucella melitensis*, *Staphylococcus saprophyticus* and *Gardnerella vaginalis*; mummified tissues, which have yielded *Helicobacter pylori*, *V. cholerae* and *Plasmodium vivax*; and frozen and formalin-fixed samples, yielding HIV and influenza.
Table 1 | Ancient pathogen genomic data recovered from archaeological or historical specimens

| Pathogen                  | Infectious disease                                  | Method of retrieval               | Number of genomes* | Biological insights                                                                                                                                                                                                 | Refs            |
|---------------------------|-----------------------------------------------------|-----------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|
| *Borrelia recurrentis*    | Relapsing fever                                     | Shotgun sequencing                | 1                  | • Isolation from 15th-century CE human remains from Norway  
  • 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  
  • 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  
  • Association with women's mortality during childbirth in the past  
  • The identified strain clusters among modern *G. vaginalis* diversity                                                                                                                                     | 46              |
| *Helicobacter pylori*     | Ulcers of the upper gastrointestinal tract  
  • Increased risk of gastric carcinoma                  | In-solution capture followed by NGS | 1                  | • Isolation from European Copper Age, 5,300-year-old mummy (Ötzi)  
  • Unadmixed strain, contrary to modern European strains, which are hybrids of two ancestral populations                                                                                               | 49              |
| *Mycobacterium leprae*    | Lepromatous leprosy                                  | Shotgun sequencing                | 27                 | • First de novo assembled ancient pathogen genome  
  • Estimated emergence >5,000 years ago  
  • European origin of leprosy in the Americas  
  • High *M. leprae* diversity in medieval Europe                                                                                                   | 23,105,201      |
| *Mycobacterium tuberculosis* | Tuberculosis                                        | Shotgun sequencing                | 19                 | • Genomes from pre-Columbian human infections show phylogenetic clustering within animal-adapted lineage present today in seals  
  • Molecular dating analysis suggests emergence of MTBC <6,000 years ago  
  • Analysis of European genomes shows past occurrence of multiple infections and suggests origin of lineage 4 during the 4th to 5th century CE                                                                 | 20,52,123       |
| *Salmonella enterica*     | Enteric (paratyphoid) fever  
  *Paratyphi C*         | Shotgun sequencing                | 11                 | • S. enterica subsp. enterica serovar Paratyphi C presence in 12th-century CE Norway  
  • 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*          | Urinary tract infections  
  *S. saprophyticus*   | Shotgun sequencing                | 1                  | • Identified in ~800-year-old human remains from Troy  
  • Association with women's mortality during childbirth in the past  
  • 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)  
  • First pathogen genome reconstructed from ancient dental calculus                                                                                                                                      | 46              |
| *Treponema pallidum*      | Syphilis (Treponema pallidum subsp. pallidum)  
  *Yaws (Treponema pallidum subsp. pertenue)*  
  *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  
  • Different Treponema subspecies (T. pallidum subsp. pallidum and subsp. pertenue) caused similar skeletal lesions usually identifiable as skeletal syphilis in infants                                                                 | 26              |
| *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  
  • The identified strain shows highest similarity with the classic pathogenic biotype O1                                                                                                                     | 55              |
| *Yersinia pestis*         | Bubonic, pneumonic and septicaemic plague           | Shotgun sequencing                | 38                 | • Bacterium affected humans as early as 5,000 years ago  
  • Both flea-adapted and non-adapted variants were present in Eurasia during the Bronze Age  
  • Causative agent of the Plague of Justinian (6th century CE)  
  • Causative agent of Black Death and persistence in Europe during the second plague pandemic (14th to 18th century CE)  
  • Possible European origin of third plague pandemic lineage                                                                                                                                             | 20,39–43        |
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)\(^{57-60}\), as well as hybridization-based enrichment techniques\(^{57,61-63}\) (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\(^{62-64}\) or when a hypothesis exists for the causative agent of an epidemic\(^{65}\). By contrast, broad laboratory-based pathogen screening in aDNA research has used microarrays for both targeted enrichment\(^{66}\) and fluorescence-based methods.

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

---

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

| Pathogen | Infectious disease | Method of retrieval | Number of genomes* | Biological insights | Refs |
|----------|-------------------|---------------------|--------------------|---------------------|------|
| **Viral pathogens** | | | | | |
| HBV | Viral hepatitis | • Shotgun sequencing  
• In-solution capture followed by NGS  
• Whole-genome PCR\(^b\) | 17 | • Identified in ancient human specimens as early as 7,000 years ago  
• Neolithic genome lineage related to contemporary strains identified in African non-human primates  
• Complex evolutionary history of HBV and identification of ancient recombination event giving rise to genotype A strains | \(^{41,44}\) |
| 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  
• HIV was introduced into the Americas from the Caribbean in the early 1970s | \(^{57}\) |
| B19V | | • Erythema infectiosum (fifth disease) in children  
• Arthropathies in adults  
• Hydrops fetalis or fetal death in pregnant women  
• 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  
• 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  
• Avian source of 1918 influenza pandemic (Spanish flu, 1918–1920)  
• 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  
• 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  
• Isolated from historical herbarium specimens  
• A unique \(P\). \(i\). \(n\). \(f\). \(i\). \(n\). \(f\). \(a\). \(n\)\(e\)\(e\)\(s\)\(s\)\(_\text{f}9,60\) genotype caused the Irish potato famine and during the 1900s became replaced by the US-1 lineage that dominated worldwide until the 1970s | \(^{55,60}\) |
| Plasmodium falciparum and Plasmodium vivax | Malaria | In-solution capture followed by NGS | 5 | • Oldest \(P\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\)\(_\text{f}9,60\) detection from southern Italy (1st to 2nd century CE)  
• \(P\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\)\(_\text{f}9,60\) and \(P\). \(v\). \(a\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\). \(_\text{f}9,60\) mitochondrial genome isolation from 20th century microscopy slides  
• Possible introduction of \(P\). \(v\). \(a\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\). \(_\text{f}9,60\) 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. *The indicated numbers include whole pathogen genomes and specimens yielding genome-wide data. †Whole-genome PCR amplicons from the studies of influenza virus\(^a\), HIV\(^a\) and HBV\(^a\) that were sequenced using capillary sequencing (Sanger method).  

---

\(^{a}\)RNA; and dried plant leaves from herbarium collections, preserving \(P\). \(i\). \(n\). \(f\). \(a\). \(n\). \(f\). \(a\). \(n\)\(_\text{f}9,60\), 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 minuscule 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).
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. 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. 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 has enabled pathogen screening via fluorescence-based detection on microarrays and via DNA enrichment approaches. The latter has been achieved, through single locus in-solution capture or through simultaneous screening for multiple pathogens using microarray-based enrichment of species-specific loci 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, 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.
deamination of cytosine most often refers to the hydrolytic removal of an amine group (NH₂) from a DNA strand. 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 5ʹ ends of aDNA fragments is often considered a criterion for authenticity. 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). 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. Consequently, the evaluation of DNA damage profiles (for instance, by using mapDamage2.0 [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.

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 or partially 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. 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.

More recently, in-solution-based capture approaches have gained popularity owing to their capacity for greater sample throughput without compromising capture efficiency; 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, even from climate zones that pose challenges to aDNA preservation. Nevertheless, deep shotgun sequencing alone has also been used for human whole-genome high-quality
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 or palaeopathological evaluation of ancient skeletal remains or analysis of historical records. 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. Similarly, historically recorded symptoms can often be misinterpreted given that past descriptions may be unspecific and do not always conform to modern medical terminology.

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, 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, 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.

Regarding bacterial pathogens, the identification of *B. recurrentis* in a 15th-century individual from Norway showed that — aside from *Y. pestis* — other vectorborne 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 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) 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, 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, viruses or mitochondrial genomes and nuclear genome data from eukaryotic microorganisms. 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. In addition, histograms of SNP allele frequencies — used to estimate the frequency of heterozygous calls in haploid organisms — can often demonstrate the effects of environmental contamination on ancient microbial data sets.

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. Among the most commonly used tools in ancient microbial genomics are MEGA, which comprises several phylogenetic methods; PhyML, RAxML and IQ-TREE, which implement maximum-likelihood approaches; MrBayes, which uses a Bayesian approach; and programs used for phylogenetic network inference, such as SplitsTree. 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. Considered alongside the oldest palaeopathological cases of leprosy dating to as early as the Copper and Bronze Age in Eurasia, these results may suggest a long history of *M. leprae* presence in this region. Moreover, the phylogenetic analysis of a 12th-century *S. enterica* genome from Europe showed its placement within the Paratyphi C lineage. Further identification of the bacterium in 16th-century colonial Mexico 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, 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 and Recombination Detection Program 4 (RDP4) have been used to infer potential recombination regions within ancient
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 (¹⁴C) 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.

bacteria28,37 and viruses43,44,53, respectively. In addition, principal component analysis (PCA) and ancient admixture component estimation using the Bayesian modelling frameworks STRUCTURE29 and fineSTRUCTURE30 on both multilocus sequence typing (MLST) and whole-genome data were recently used for population assignment of a 5,300-year-old *H. pylori* genome31. 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* used the program TREE-PUZZLE32, 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 anthropogenic 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 lions33. 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 contact34. 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 primates35,36, 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 Americas37, reinforcing the hypothesis that humans passed the disease to the nine-banded armadillo, the most common reservoir for this disease in the New World38.

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 costs39 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) database40 and the recently introduced EnteroBase41) and the development of robust phylogenetic frameworks that can accommodate genome-wide data from >100,000 strains (for example, GrapeTree42) 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.

**Infering 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 set43. For such analyses, the most widely used program is the Bayesian statistical framework BEAST44,45.

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 ago46. 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 Peru47, a result that was further
corroborated by the incorporation of 18th-century European MTBC genomes in the dating analysis\textsuperscript{20,23,25,26}.

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\textsuperscript{24}, whereas the latter is used for estimation of a correlation coefficient (r) and coefficient of determination (R\textsuperscript{2}) by relating the tip date of each taxon to its SNP distance from the tree root (using, for example, the program TempEst\textsuperscript{23}). 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\textsuperscript{28} can be statistically tested, for example, using MEGA9\textsuperscript{29} or marginal likelihood estimations\textsuperscript{127,128}, and applied. If branches are affected by differences in their evolutionary rates, a relaxed clock\textsuperscript{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\textsuperscript{30,31}. 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\textsuperscript{30,32}. Similar limitations can arise when the evolutionary history of a microorganism is vastly affected by recombination, as observed for HBV\textsuperscript{33,34}, 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\textsuperscript{35}.

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\textsuperscript{127,128}). Currently, the most widely used models for estimating dates of divergence are the coalescent constant size\textsuperscript{37}, which assumes a continuous population size history — and is unrealistic for epidemic pathogens — and the coalescent skyline\textsuperscript{132}, which can estimate effective population size (N\textsubscript{e}) changes over time. Moreover, the birth–death demographic model\textsuperscript{33,134}, 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\textsuperscript{135}. 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\textsuperscript{136}.

**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\textsuperscript{20,23,25–38} (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\textsuperscript{139,140}. 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\textsuperscript{141}. 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\textsuperscript{142}. 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 *Y. 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\textsuperscript{138}, and pMT1, which is involved in flea colonization\textsuperscript{139,140}, as well as by chromosomal gene pseudogenization or loss throughout its evolutionary history\textsuperscript{142}.

*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\textsuperscript{142–144}. Its global distribution includes rodent species\textsuperscript{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\textsuperscript{139,143,147,148}. *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\textsuperscript{142,143,146,147}. Notably, recent modelling inferences suggest important roles for ectoparasites such as body lice and human fleas in its propagation during human epidemics\textsuperscript{148}. 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
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 and facilitates colonization of the arthropod midgut. 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) and the urease gene ureD, 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. This continuous transmission cycle among fleas and rodents, also called the enzootic phase (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. 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 (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. Bubonic plague is the most common form of the disease and can cause up to 60% mortality when left untreated. 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.
Based on aDNA data (13,000–79,000 years before present (ybp))\(^{33,34}\). Nevertheless, \textit{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 \textit{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 \textit{ymt}, whose absence has been associated with an inability for flea midgut colonization in \textit{Y. pestis}\(^{141}\). In addition, these strains possess the active forms of the \textit{rcsA}, \textit{PDE3}, \textit{PDE2} and \textit{ureD} genes, which suggests an impaired ability towards biofilm formation and blockage of the flea’s proventriculus\(^{41,46}\). Finally, they possess an active flagellin gene (\textit{flhD}), which is present as a pseudogene in all other \textit{Y. pestis}, as it is a potent inducer of the innate immune response of the host\(^{138}\). As a result, during its initial evolutionary stages, \textit{Y. pestis} may have been unable to efficiently transmit via a flea vector. Flea-borne transmission of \textit{Y. pestis} is a known prerequisite for bubonic plague development\(^{161}\); hence, it has been suggested that this disease phenotype was not present during prehistoric times\(^{33,136}\). 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 \textit{Y. pestis} LNBA lineage was shown to mirror human movements during that time\(^{4}\) 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 \textit{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 \textit{Y. pestis} genomic

---

**Fig. 5 | \textit{Yersinia pestis} ecology and transmission cycle.** A simplified version of the \textit{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 \textit{Y. pestis}. Plague can manifest in humans in the bubonic, pneumonic and septicemic forms. Pneumonic plague is the only form that can result in airborne transmission between humans.
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. 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, 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 and B19V (TABLE 1), were circulating during the same time periods. Further pathogen screening coupled with a temporal assessment of human immune-associated genomic variants 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 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) 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.

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. In addition, historical accounts of the disease seemed to correlate with the supposed distinct geographical distributions of these biovars, and their phylogenetic relationships, as inferred from MLST data, reinforced the hypothesis that each was responsible for a single pandemic. By contrast, later studies identified additional, atypical biovars, and more robust phylogenetic analysis suggested that phylogeography does not correlate clearly with the phenotypic distinctions described between these bacterial populations.

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. 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. More recent investigations have revealed previously uncharacterized genetic diversity in the geographical distributions of these biovars, and their phylogenetic relationships, as inferred from MLST data, reinforced the hypothesis that each was responsible for a single pandemic. By contrast, later studies identified additional, atypical biovars, and more robust phylogenetic analysis suggested that phylogeography does not correlate clearly with the phenotypic distinctions described between these bacterial populations.

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. 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. The analysis of aDNA from historical epidemic 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. 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. More recently, full characterization and authentication of the bacterium were achieved using plasmid and whole-genome enrichment coupled with NGS.

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. Two recent studies have reconstructed 6th-century *Y. pestis* genomes from southern Germany, 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, 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, which shares a common ancestor with the Justinianic-plague lineage. However, given the >300-year age difference between these strains, as well as the aforementioned East Asian sampling bias of modern *Y. pestis* data, 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.
Death: Y. pestis strain identified in 14th-century western, northern and southern Europe [9,10,13,14], suggesting that they likely arose locally. The first lineage survives today and gave rise to modern branch 1 strains [9,10,13,14] (which are associated with the third plague pandemic), suggesting the European Black Death as a source for modern-day epidemics [9]. The second lineage has not been identified among present-day diversity and currently encompasses strains from 16th-century Germany [9,10] 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 [10].

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 [15,16]. The reasons for its disappearance are unknown, although numerous hypotheses have been put forward [17,18], including a change in domestic rodent populations in Europe; namely, the replacement of the black rat, Rattus rattus, by the brown rat, Rattus norvegicus [19]; an acquired plague immunity among humans and/or rodents [20,21] (although this hypothesis requires an update to accommodate the recent identification of Y. pestis in Europe 5,000 years ago [22,23], and the involvement of the bacterium in the first plague pandemic [24]), 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 [25,26]; and the potential disruption of the European wild rodent ecological niche owing to habitat loss and industrialization starting in 1700 CE [27]. 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. E. & 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., Széminés, U. & Nerlich, A. G. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J. Med. Microbiol.* 50, 355–366 (2001).
8. Drancourt, M., Abouhashem, 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 septicaemia. *Proc. Natl Acad. Sci. USA* 95, 12637–12640 (1998).
9. Spiegelman, 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, 1939–1945 (1989).
11. Sawyer, S., Krause, J., Guschnick, K., Savolainen, V. & Pääbo, 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).
13. The study provides a quantitative description of aDNA-associated patterns of nucleotide misincorporation and fragmentation that are currently used as primary authentication criteria.
14. Cooper, A. & Poinar, H. N. Ancient DNA: do it right or not at all. *Science* 289, 1139–1139 (2000).
15. 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).
16. Shapiro, B., Rambaut, A. & Gilbert, M. T. P. No proof that typhoid caused the Plague of Athens (a reply to Papapapagiokas et al.). *Int. J. Infect. Dis.* 10, 334–335 (2006).
17. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376 (2005).
18. 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).
19. Green, R. E. et al. The Neandertal genome and ancient DNA authenticity. *EMBO J.* 28, 2494–2502 (2009).
20. Key, R., Poth, C., Krause, J., Herbig, A. & Box, K. I. Merging metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends Genet.* 33, 508–520 (2017).
21. Bos, K. I. et al. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506–510 (2011).
22. The study describes the first whole-genome sequence of an ancient bacterial pathogen through the use of high-throughput sequencing.
Duggan, A. T. et al. 17th century variola virus reveals timing of virulence determinant acquisition during The study describes in Eurasia 5,000 years ago. Mol. Biol. Evol. 33, 2911–2923 (2016).

Seifert, L. et al. Genotyping Y. pestis in historical plague evidence: identification of Y. pestis in Europe from the 14th to the 17th century. PLOS ONE 11, e0145194 (2016).

Harbeck, M. et al. Ancient Y. pestis DNA from skeletal remains of the 6th century AD reveals insights into Justian Plague. PLOS Pathog. 9, e100349 (2013).

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

I. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic screening of ancient DNA. Nucleic Acids Res. 34, D589–D595 (2010).

M. Morelli, G. et al. A high-coverage genome sequence of Y. pestis from a 16th century mummy. Nature 437, 889 (2005).

J. Yabu da, K. et al. The rise and fall of the Phythophobia infestans lineage that triggered the Irish potato famine. eLife 2, e00731 (2013).

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

H. Kamps, K. M. et al. Screening ancient tuberculosis with aPCR: clinical utilities. Phil. Trans. R. Soc. B 370, 2015032622 (2015).

Worobey, M. et al. 1970s and ‘patient 0’HIV-1 ancestor in a 15th century Nigerian grave. Science 340, 739–743 (2013).

M. Müller, R., Roberts, C. A. & Brown, T. A. Complications of Mycobacterium leprae partial leprosy in medieval Europeans. Nature 519, 214–218 (2015).

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.

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

R. C. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. 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.

J. Prüfer, K. et al. The complete genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757–762 (2010).

J. Prüfer, K. et al. The complete genome sequence of an extinct Pleistocene. Nature 505, 43–49 (2014).

J. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

E. Morell, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1104–1106 (2010).

J. 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.

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

C. C. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. 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.

J. Prüfer, K. et al. The complete genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757–762 (2010).

J. Prüfer, K. et al. The complete genome sequence of an extinct Pleistocene. Nature 505, 43–49 (2014).

J. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1104–1106 (2010).

J. 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.

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

C. C. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. 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.

J. Prüfer, K. et al. The complete genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757–762 (2010).

J. Prüfer, K. et al. The complete genome sequence of an extinct Pleistocene. Nature 505, 43–49 (2014).

J. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1104–1106 (2010).

J. 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.

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

C. C. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1104–1106 (2010).

J. 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.

J. Prüfer, K. et al. The complete genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757–762 (2010).

J. Prüfer, K. et al. The complete genome sequence of an extinct Pleistocene. Nature 505, 43–49 (2014).

J. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 224–228 (2012).

G. Morelli, G. et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat. Genet. 42, 1104–1106 (2010).

J. 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.
Ferreira, M. A. & Suchard, M. A. Bayesian analysis. Rambaut, A., Lam, T. T., Max Carvalho, L. & Baele, G. et al. Improving the accuracy of genome assembly and multilocus Salmonella Achtman, M. A genomic overview of the population (GSP).

Can. J. Stat. of elapsed times in continuous- time Markov chains. Pybus, O. G. Exploring the temporal structure of evolutionary analysis by sampling trees.

PLOS Comput. Biol. tuberculosis in a mummy. The study presents a comprehensive

PLOS Comput. Biol. 58:513–525 (2013).

Chain, P. S. et al. Insights into the evolution of virulence in mice.

Yersinia pestis, the cause of plague, while accommodating phylogenetic uncertainty. Rambaut, A. & Fetherston, J. D. Verapamil—estriol agent of plague. Clin. Microbiol. Rev. 10, 554–557 (2017).

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

Miranda, S. A. N. The role of Yersinia pestis from Genomics to Function (eds Perry, R. D. & Fetherston, J. D.) Vol. 603 298–311 (Springer, 2007).

Hinnebusch, B. J. Choukroun, S. & Sun, V. C. Ecological opportunity, evolution, and the emergence of flea-borne plague. Infect. Immun. 84, 1952–1960 (2016).

Allentoft, M. E. et al. Population genomics of the Bronze Age Eurasia. Nature 466, 167–172 (2010).

virus. The same data set was later used for pathogen screening and could show the presence of Y. pestis, HBV and B19V in those populations.

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, 1185–1192 (2015).

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

Harper, K. The Fate of Rome: Climate, Disease, and the End of an Empire (Springer, 2017).

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

Politzer, R. Plague (eds Pachowicz, M. & Moskvitch, A.) in Monographs Series, no. 22 (World Health Organization, 1954).

Desvignat, R. Varieties of the Espece Pasteurella pestis: nouvelle hypothèse [French]. Bull. World Health Organ. 4, 247–265 (1951).

Zhong, D. et al. Genetics of metabolic variations in Yersinia pestis. Front. Microbiol. 2, 5 (2012).

Kutuvary, 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, 308 (2018).

Eroschenko, G. A. et al. Y. pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. PLOS ONE 12, e0182005 (2017).

Kulisichkina, A. A. et al. Nineteen whole-genome assemblies of Yersinia pestis subspecies microtus, including representative strains isolated in the caucasus region. Genomics Data 21, 100158 (2020).

Zhang, E. et al. Genomic analysis of Y. pestis strains isolated in the caucasus region. Genomics Data 3, e10350–15 (2015).

Schafer, A. J. et al. Temporal phylogeography of Yersinia pestis in Madagascar: insights into the long-term maintenance of plague. PLOS Negl. Trop. Dis. 11, e0005387 (2017).

Drancourt, M. et al. Y. pestis Orientalis in remains of ancient plague patients. Emerg. Infect. Dis. 13, 352 (2007).

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

McGonnell, M. Tracking mass death during the fall of empire’s frontier. J. Roman Archaeol. 28, 325–357 (2015).

Giraud, U., Gruaz, C., Esper, J., Tjegl, W. & McMichael, A. J. Digitizing historical plague. Clin. Infect. Dis. 55, 1586–1588 (2012).

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

Schmid, B. V. et al. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. Proc. Natl Acad. Sci. USA 112, 423–428 (2015).

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

Appleby, A. B. The disappearance of plague: a continuing puzzle. Hist. Rev. 35, 161–173 (1980).
182. DeWitte, S. N. Mortality risk and survival in the aftermath of the medieval Black Death. PLOS ONE 9, e96515 (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 Horb. 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 str. w1118; iso-2; polymorphisms, SnpEff: SNPs in the genome of (2017).
187. Kislichkina, A. A. et al. Nine whole-genome assemblies of Yersinia pestis subsp. micrcotus bv. Altaica strains isolated from the altai mountain natural plague focus (No. 56) in Russia. Genome Announc. 6, e01440–17 (2018).
188. 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).
189. 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).
190. Parkhill, J. et al. Genome sequence of Yersinia pestis, the causative agent of plague. Nature 413, 523 (2001).
191. 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).
192. Gibbons, H. S. et al. Comparative genomics of 2009 seasonal plague (Yersinia pestis) in New Mexico. PLOS ONE 7, e51604 (2012).
193. Shen, X. et al. Complete genome sequences of Yersinia pestis from natural foci in China. J. Bacteriol. 192, 3551–3552 (2010).
194. Touchman, J. W. et al. A North American Yersinia pestis draft genome sequence: SNPs and phylogenetic analysis. PLOS ONE 2, e220 (2007).
195. Malek, M. A. et al. Yersinia pestis halotolerance illuminates plague reservoirs. Sci. Rep. 7, 40022 (2017).
196. Mendum, T. A. et al. Mycobacterium leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC Genomics 15, 270 (2014).
197. Tumpey, T. M. et al. Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science 310, 77–80 (2005).

Acknowledgements
The authors thank C. Warinner for her valuable comments to the manuscript and M. Keller for his contributions in assembling comprehensive meta-information for the Y. pestis modern genomic data set. In addition, the authors thank all members of the Molecular Paleopathology and Computational Pathogenomics groups at the Max Planck Institute for the Science of Human History for insightful discussions during meetings. Moreover, they are grateful to M. O’Reilly, H. Shell and R. Barquera for extensive assistance with the graphics. This work was supported by the Max Planck Society.

Author contributions
M.A.S. researched the literature and wrote the article. All authors provided substantial contributions to discussions of the content and reviewed and/or edited the manuscript.

Competing interests
The authors declare no competing interests.

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Genetics thanks E. Willerslev and other anonymous reviewer(s) for their contribution to the peer review of this work.