Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Current Biology

CellPress

Review

The Recovery, Interpretation and Use of Ancient Pathogen Genomes

Sebastián Duchêne[1,*], Simon Y.W. Ho[2], Ann G. Carmichael[3], Edward C. Holmes[4,*], and Hendrik Poinar[5,6,7,*]
[1]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC 3000, Australia
[2]School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia
[3]History Department, Indiana University, Bloomington, IN, USA
[4]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, University of Sydney, Sydney, NSW 2006, Australia
[5]McMaster Ancient DNA Centre, Departments of Anthropology and Biochemistry, McMaster University, 1280 Main St. W., Hamilton, ON L8S 4L9, Canada
[6]Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, 1280 Main St. W., Hamilton, ON L8S 4L8, Canada
[7]Humans and the Microbiome Program, Canadian Institute for Advanced Research, Toronto, Canada
*Correspondence: sebastian.duchene@unimelb.edu.au (S.D.), edward.holmes@sydney.edu.au (E.C.H.), poinarh@mcmaster.ca (H.P.)
https://doi.org/10.1016/j.cub.2020.08.081

SUMMARY

The ability to sequence genomes from ancient biological material has provided a rich source of information for evolutionary biology and engaged considerable public interest. Although most studies of ancient genomes have focused on vertebrates, particularly archaic humans, newer technologies allow the capture of microbial pathogens and microbiomes from ancient and historical human and non-human remains. This coming of age has been made possible by techniques that allow the preferential capture and amplification of discrete genomes from a background of predominantly host and environmental DNA. There are now near-complete ancient genome sequences for three pathogens of considerable historical interest — pre-modern bubonic plague (*Yersinia pestis*), smallpox (*Variola virus*) and cholera (*Vibrio cholerae*) — and for three equally important endemic human disease agents — *Mycobacterium tuberculosis* (tuberculosis), *Mycobacterium leprae* (leprosy) and *Treponema pallidum pallidum* (syphilis). Genomic data from these pathogens have extended earlier work by paleopathologists. There have been efforts to sequence the genomes of additional ancient pathogens, with the potential to broaden our understanding of the infectious disease burden common to past populations from the Bronze Age to the early 20th century. In this review we describe the state-of-the-art of this rapidly developing field, highlight the contributions of ancient pathogen genomics to multidisciplinary endeavors and describe some of the limitations in resolving questions about the emergence and long-term evolution of pathogens.

Introduction

The evolutionary and epidemiological history of human pathogens reflects a combination of short- and long-term interactions among biological, ecological and cultural factors. Although these processes have been studied using a variety of approaches, work over the past 20 years has shown how studies of ancient DNA and ancient RNA can provide important insights into the origins and history of infectious disease. For example, ancient DNA has led to a new understanding of endemic human diseases that span millennia, such as malaria, leprosy and tuberculosis [1–3], as well as sporadic and devastating pandemic diseases, such as the recurrences of plague across Eurasia [4]. Reconstructing the evolutionary history of these pathogens with the help of ancient DNA has illuminated novel aspects of their evolution and epidemiology and has provided unexpected insights into the nature of disease-associated microbes [5–13].

By the early 1990s, ancient DNA work on tuberculosis and leprosy had stoked optimism that genomic techniques could provide informative new data on the origin and evolution of infections still common today. Nevertheless, most human pathogens remained undetected in ancient samples until novel high-throughput sequencing and targeted enrichment offered enhanced resolution. For example, this genomic-scale work has led to revised narratives on the temporal-spatial emergence and transmission of *M. tuberculosis* and *M. leprae*, generated indisputable evidence of the bacterium responsible for yaws (*Treponema palladium pertenue*) being present in Western Europe before the journeys of Columbus and revealed the global dispersal of ubiquitous human pathogens, including hepatitis B virus (Box 1) and *Helicobacter pylori*. In combination, this work will continue to revise our understanding of the history of infectious disease [4,13–15].

The causes of differential mortality within populations had previously received little more than rudimentary attention in most histories of pre-modern disease. Ancient DNA provided a new perspective, by showing endemic infectious diseases increased the risk of dying during times of major epidemics such as the

Box 1. Genetics does not meet history: Hepatitis B virus.

In the 1950s, Baruch S. Blumberg, who was interested in human variation in susceptibility and resistance to disease, began a global-scale collection of human serum. He identified a specific antigen present in the blood of hepatitis victims, called the ''Australia antigen'' because it was also present in an indigenous Australian. In the 1970s, David S. Dane identified the presence of virus-like particles with antigen-positive serum, leading to the isolation of hepatitis B virus (HBV) and now the identification of 10 genotypes (A through J), with geographic distributions differing among genomes (for example, some genotypes are restricted to the Americas). Until recently the historical evidence for HBV infection in humans spanned less than 100 years, although the global distribution and high prevalence of the virus in many populations suggested that it might have been associated with humans since their dispersal out of Africa [125].

Molecular clock analyses of contemporary genomes were unable to paint a clear picture of HBV origins, nor the timescale over which it has occurred. However, the integration of ancient DNA shows that the virus evolves slowly in the long term: for example, there was no evidence for temporal signal in the data even with the addition of ancient genomes recovered from 16th century child mummies from Italy and Korea [9,126]. Consequently, some researchers have attempted to calibrate the evolutionary timescale of HBV by assuming codivergence of the viral genotypes with their human host populations [111,127].

Analyses of HBV genomes from the two child mummies showed that the genomes were nested within the contemporary diversity of the virus in these locations, indicating that the present genotype distribution is at least 400 years old. However, some ancient HBV genomes reconstructed from Bronze Age (~5000 years ago) and Iron Age (~1500 years ago) human remains from Eastern and Western Europe are actually closer to lineages circulating in chimpanzees and gorillas today [128,129] (Figure 3). Although the extent of temporal signal is still uncertain even with the inclusion of the Bronze Age genomes (Figure 3C), the phylogenetic position of the ancient samples indicates that HBV has had a long association with human populations, spanning at least seven thousand years, and that its present distribution can largely be explained by normal and/or forced human migration [130]. Nevertheless, the exact patterns of migration, including when and how the virus arrived in the Americas, remain unclear.

Black Death [16,17]. For instance, *Salmonella enterica* contributed to catastrophic population collapse in Mesoamerica during the 16th century European conquest, but also was part of the background disease landscape in 13th century Norway [11,18,19]. Similarly, pathogens fortuitously retrieved from individual skeletal remains provide important evidence related to intervals of known, but poorly understood, mortality crises. For example, historical evidence of the toll that typically survivable infections could cause in crisis circumstances is illustrated by the Great Irish Famine (Box 2). Incidental and unexpected findings are also becoming increasingly routine with ancient pathogen genomics, enriching our biological and cultural understanding of the past. As a case in point, some of the best and more surprising sources of pathogen ancient DNA are calcified abscesses and kidney, urinary and placental stones [12,20]. The analysis of vaginal flora isolated from a calcified abscess in the skeletal remains of a female from 13th century CE Troy reminds us that maternal sepsis was, and still is, a common cause of death among adult females.

Despite the evident advances stemming from ancient DNA, it is also clear that increased efforts to include additional disciplines (e.g., historians and ethicists) and stakeholders (e.g., descendants, relatives or community relations and museums) are needed for the true value of ancient pathogen genomes to be properly contextualized [21]. In particular, an immediate challenge that we will discuss below is how to place genetic evidence into a more inclusive and expansive historical context, augmented by climatological, evolutionary and archaeological frameworks [22].

Ancient DNA and RNA Degradation and Preservation

A new understanding of past human disease experience is built on the preservation and extraction of unstable and damage-prone nucleic acids. Ancient DNA and ancient RNA are relative terms used to characterize the damaged nucleic acids isolated from variously aged samples. DNA damage begins immediately *post mortem* during necrosis, via host enzymes and microbiomes, as well as via chemical pathways such as hydrolysis, alkylation, condensation and oxidation [23,24]. Once the body is buried, stored or treated (the rate of which matters in each case), further damage to DNA occurs via exogenous microbial and fungal contaminants which leach into the remains over time (Figure 1). More broadly, the conditions under which DNA is degraded or preserved can vary dramatically by the regional geology, climate and weather, by proximate environmental conditions (cave *versus* open-air sites, burial depth and humidity) and at the microscopic level (skeletal elements, mineral content and exposure) [23–25]. Given the tremendous variation in preservation potential, there is no simple correlation between the level of DNA damage and the age of a sample or its origin. Abiotic environmental constraints limit access to ancient DNA from historically relevant contexts (Africa, the Middle East and Central and South America), but exceptions, such as cave sites, may prove successful.

Ancient DNA molecules are short, usually between 30 to 60 base pairs long, the result of hydrolytic damage to both the phosphodiester bond (the weakest link in DNA) and the glycosidic bond [26,27], as well as oxidative damage to the sugar rings and pyrimidines [27,28]. Genome stretches break into increasingly shorter DNA molecules, in a manner that is largely dependent on the availability of free water and the movement of reactants [23]. Double-stranded fragments often contain short single-stranded overhangs that are more susceptible to modification. Ancient DNA termini contain modified phosphates that are likely to be inaccessible to current library-preparation methods [28]. In addition, amine-containing nucleobases, such as cytosines (and adenine and guanine to a lesser extent), found in these single-stranded overhangs are subject to hydrolytic

Box 2. Genetics meets history: *Phytophthora infestans*.

The plant pathogen *Phytophthora infestans*, a eukaryotic oomycete that causes potato late blight, arrived in Europe by 1845 and destroyed potato crops across much of the continent [131]. The resultant Great Irish Famine became a demographic catastrophe because a quarter of Ireland's people subsisted on potatoes and a great many others relied on the potato as a primary staple. Over a million persons emigrated, leaving the net demographic decline near 30% of Ireland's population (worse than what Ireland suffered during the Black Death), though relatively few people died from outright starvation.

Demographic and epidemiological analysis of the causes of death in this subsistence crisis [132] shows patterns common in historical mortality crises. Cause-of-death reports show that many different infectious diseases circulated and were made more lethal by progressive undernutrition and malnutrition. Still-able hunger victims migrated to towns and cities in search of food or work. Tickborne typhus (*Rickettsia prowaszecki*) and relapsing fever (*Borrelia recurrentis*) in particular added to the poorly defined 'famine fevers' that were the leading cause of death and illness. Deaths from diarrhea and dysentery during these years were amplified by the second wave of pandemic cholera, making it difficult to distinguish the burden of each disease. Transatlantic outbreaks of virulent streptococcal infections might have contributed to deaths from 'rheumatism', but would not have been distinctive among rashes from scurvy or the various epidemic fevers.

Botanists at the time described the pathogen and its effects on potato plants. Ancient genomic DNA recovered from preserved leaves in several different European herbaria illustrates some of the uses and limitations of recent research [133–137]. Phylogenetic analyses of complete nuclear and mitochondrial genomes from *P. infestans* have painted a complex picture of host switching and hybridization with *P. andina*, an emerging pathogen of solanaceous crops [135]. The genomic evidence supports an origin for *P. infestans* in Mexico or the South American Andes [134–137]. The HERB-1 lineage of the pathogen, which was responsible for the Irish potato famine, had either multiple introductions into Europe or a single introduction of multiple haplotypes in the 19th century [135]. Although the HERB-1 lineage was believed to have gone extinct [137], analyses of strains from the Americas have revealed that it persists today in Central and South America [134].

deamination, which results in uracil. Uracils that are subsequently copied as thymines in amplification reactions lead to C-to-T and G-to-A miscoding lesions, the tell-tale sign of ancient DNA authenticity [24,29,30].

Ancient DNA isolated from skeletal remains is a complex mix of endogenous (target) with exogenous (non-target) DNAs that have worked their way into the remains over their burial history. As endogenous DNA of the sample degrades and leaches into the sedimentary environment, exogenous DNAs from the environment do the opposite. The equilibrium between these DNAs is guided mostly by the availability and movement of free water, bacteria, biofilms and fungal hyphae. Finally, exogenous DNAs can infiltrate the sample or extracts during or after excavation, handling, storage and laboratory-based methodologies or activities [31]. Between 60% and 90% of ancient DNA is lost during extraction, library preparation and the numerous purification steps that render the extracts free of their often extensive, enzyme-inhibiting coeluates [24,32]. The target constituent of ancient DNA extracts typically ranges between 1% and 10%, with a detectable pathogen fraction usually far less than 0.5% of the total metagenome [33–35], although exceptions have been found in calcified nodules, ear ossicles and permafrost samples [12,20,36].

The even faster degradation of RNA molecules places limitations on evolutionary studies of pathogens with RNA genomes, such as influenza virus, measles virus, HIV and yellow fever virus. The 2' hydroxyl group of RNA increases the strength of the glycosidic bond, slowing depurination but weakening the phosphodiester bond. The nearly ubiquitous presence of RNases explains the further rapid degradation of RNA. Little work has been done on the preservation and degradation of RNA from ancient samples, although recent studies suggest that RNA can survive under favorable conditions for millennia [37,38]. On the other hand, formalin fixation, particularly problematic for DNA, appears to preserve RNA and this has led to exciting recoveries of RNA viruses from stored archival tissues [39–42].

Pathogen Fortitude

Differential preservation of pathogens plays a key role in their survival and thus their direct association with some historically important epidemics and infections, but not others. Bacterial endospores fall on one end of the spectrum, with species such as *Bacillus anthracis* and *Clostridium difficile* the most suited for long-term DNA preservation [43]. Endospores have been revived after decades of dormancy, yet they have not been detected nor are they being sought from ancient remains.

Most bacteria have a single wall and membrane composed entirely of peptidoglycan. Gram-positive bacteria have a two- to ten-fold thicker cell wall than Gram-negative bacteria, which may play a role in the longer preservation of their DNA. This is unfortunate, as Gram-negative bacteria account for the bulk of pathogenic bacteria. *Mycobacterium tuberculosis and M. leprae* (both Gram-positive) produce thick lipid exudates or nodules that are likely to help repel water and enhance preservation, which may account for the reduced signal of deamination that was seen in an ancient *M. leprae* genome [44].

Treponema pallidum pallidum and *pertenue* (causative agents of syphilis and yaws, respectively) sit at the other end of the spectrum because they lack an outer cell membrane altogether. This feature, coupled with the low pathogenic load of *Treponema* in late stages of infection [45], largely precluded its identification in ancient remains [45–47] until the recent application of high-throughput sequencing methods [15,17,48].

Some viruses have been shown to survive and still be infectious after freeze-drying and being at room temperature for decades, but how this correlates with long-term preservation of their nucleic acids in the geological record remains unknown [49]. Both Gram-negative bacteria and RNA viruses are

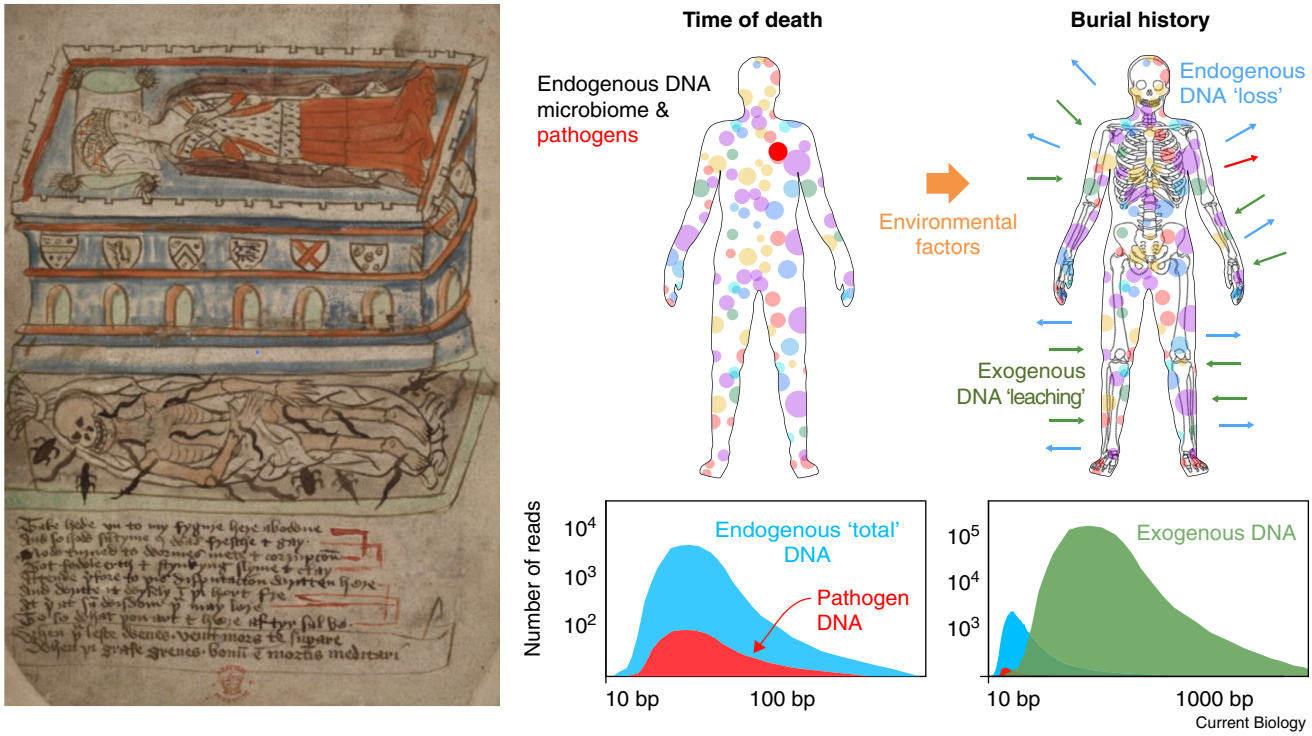


Figure 1. Degradation and preservation of ancient DNA.
Degradation and preservation of host DNA (including the microbiome and circulating pathogens) occurs rapidly after death. DNA degrades in an environmentally informed manner, where water, temperature, pH, microbial soil content and other factors shorten the remaining endogenous DNA but most is lost. At the same time, environmental (exogenous) DNA swamps out the original signal, with fragment-length distributions that both overlap and extend beyond the ancient DNA in the sample. Medieval peoples were well acquainted with death as the 15th century manuscript illustration on the left suggests (image: © British Library Board: Add MS 37049). Here, a recently buried noblewoman (endogenous) sits atop the worms and other vermin (exogenous), who will soon set upon her corpse. The image accompanies a Middle English poem relating a discussion between the decaying woman and eager worms [138].

particularly prone to rapid degradation, yet are the causative agents of many infectious diseases; they remain underrepresented in current ancient DNA studies. Further clarification on how the physical characteristics of some bacteria and viruses act to enhance or detract from nucleic acid preservation, along with improvements to ancient DNA methodologies, may help in their further targeting, detection and reconstruction.

Pathogen DNA Detection, Skeletal Lesions and Disease Identification

As the individual response to disease is complex [14,50,51], the presence of pathological lesions can be an unreliable way to sample ancient pathogen DNA. It may be better to sample from the population at large, or from a locality where an infectious agent is presumed present, such as a hospital site or mass grave. Only a handful of ancient pathogen DNA studies have relied on skeletal indicators as a sampling strategy and almost all revolve around where skeletal evidence strongly, but not solely, suggests a diagnosis of tuberculosis or leprosy [1,2,52].

In general, ancient pathogen DNA detection strongly depends on many important factors: host susceptibility to a particular infection; level of pathogen in the blood or tissue (i.e., bacteremia, viremia or parasitemia); differential immune responses; presence and level of infection at time of death; nutritional state (result of famine or social status); and prior or ongoing co-infection or co-morbidity. Although a presumed pathogen may be associated with a well-dated burial assemblage (e.g., *Y. pestis* at East Smithfield cemetery, London, UK), the level of infection at death and the burial conditions will impact the relative abundance and detectability of ancient pathogen DNA within each skeletal element. Because yields of pathogen ancient DNA can be low even in optimal sites for recovery, there are ethical ramifications that need to be carefully considered prior to beginning the work [21]. Relatively little attention has thus far been paid to the sheer bulk of finite, irreplaceable resources required in obtaining the genome sequences of ancient pathogens.

The archaeological record comprises, in decreasing order of frequency, skeletal elements (bones and teeth), mummified soft tissue, hair and human-associated trace fossils (e.g., faeces, abscesses and associated sediment). The historical period includes archival samples stored in medical museums [42,53,54]. Microbial DNA has now been isolated from almost all of these sources [55]. The successful recovery of pathogen ancient DNA depends on taking into account disease pathogenesis (i.e., localization or dissemination of microbes throughout infection and immunity responses) to guide selection of the most appropriate sampling strategy.

Pathogens that cause infections confined to soft tissue usually cannot be recovered from skeletal material, as they cause varied

clinical manifestations that primarily affect the intestinal system (e.g., inflammation, vomiting, fever or diarrhea) [56]. For these pathogens, acquiring ancient DNA requires soft-tissue samples, faecal remains, gut-associated sediment samples or archival medical or histologically preserved samples [57]. In exceptional cases, enteric pathogens can become blood borne, which might explain the recovery of *Salmonella enterica* serovar Paratyphi C (causing bacterial enteric fever) from teeth of victims from a 16th century Mesoamerican epidemic [18,58] as well as from 13th century Norway [11,19]. Hematogenous infections, such as bubonic plague, are more likely to be detectable in the blood-fed root or pulp cavity of teeth than in bones.

Commensal microorganisms, which form part of the human microbiome, inhabit diverse niches within the human body and can become opportunistic pathogens [59]. Reconstructing commensal pathogen genomes and our ancient microbiomes will clearly provide a more nuanced comprehension of health status and disease susceptibility in the past. Recent work on reconstructing these metagenomes has largely focused on using the constituents of dental calculus (calcified dental plaque) from human remains. These studies have shown the shifting composition of the oral microbiota during major transitions — such as from hunting and gathering to farming ~10,000 years ago [60] — as well as the antiquity of putative antibiotic resistance genes associated with the oral microbiota [61]. The presence, preservation and ultimate detection of ancient pathogen DNA or RNA depends on many complicating factors that include the epidemiology of the disease, the biology of pathogen and its load at time of death. These factors coupled with the fickle nature of preservation make their isolation from ancient remains both surprising and fortuitous.

Acquiring Ancient Pathogen Genomes and Microbiomes

Until 2011, pathogen detection from archaeological remains was almost entirely based on polymerase chain reaction (PCR) [24]. While PCR benefits from high sensitivity (theoretically able to detect single molecules) and specificity, most amplicons are long (~100–200 base pairs) and target a tiny fraction of the available DNA (Figure 1). When adding the complications of preservation and pathogenesis, it is unsurprising that many attempts at detecting pathogens have met with negative results. Authenticating PCR-based studies of pathogen DNA requires a number of experimental controls and replication [24]. Discriminating true human pathogens from closely related environmental relatives can be complicated; it is particularly problematic with the use of repetitive elements for the identification of *Mycobacterium* species, because these mobile elements are similar to those in other mycobacterial agents or extraneous soil microorganisms. Today PCR assays are used as a quick and easy but resource-costly way to prioritize the best libraries from a large pool of potential samples for a more targeted or deep shotgun sequencing.

With the advent of high-throughput sequencing came the move from the simple identification of pathogens at a certain place and time to genome-level reconstruction and characterization, enabling a shift to hypothesis-driven research carried out within evolutionary frameworks [62–65]. The methodological tweaks to DNA extractions and novel library preparations designed to target short, damaged and sometimes single-stranded DNA have enhanced access to the target component of ancient extracts (Box 3) [66,67]. High-throughput sequencing methods require conversion (and therefore loss) to platform-specific DNA libraries. This can involve end-repair and polymerase fill-in reactions, which will occur on only a small proportion of actual surviving and less damaged, repairable DNA molecules (those containing appropriate 5' and 3' termini) [68]. Single-stranded library preparations make the most theoretical sense because they mitigate some of the many problems associated with ancient damaged DNAs [67,69].

Pathogen-screening workflows include a shallow shotgun-sequencing approach (i.e., 1–2 million DNA reads), followed by various bioinformatic pipelines to ascertain, at first, the likelihood of pathogen-specific DNA markers. There is a growing need for new algorithms to confirm the presence of a pathogen in ancient remains, given that individual strains of microorganism contain few genomic regions that are specific at taxonomic levels and that DNA is damaged and at low quantities. Various methods exist for taxonomic binning, sequence alignment and visualization, such as MEGAN [70], HOPS [71], BLAST [72] and KRAKEN [73]. For a good overview of the applicability of these and other methods to ancient DNA, see [74].

As pathogen-specific endogenous molecules will usually be at extremely low abundance or undetectable at shallow sequencing depths, costs can quickly become prohibitive to most researchers. Shotgun sequencing as a means for the recovery of full genomes is only possible with exceptionally preserved specimens, where endogenous pathogen DNA content is high [12,20]. An alternative way to identify the few pathogen reads in a DNA extract is through the use of pathogen-enrichment arrays designed to enrich short genome-specific regions, thereby increasing the proportion of the target molecules [35,75,76]. Probe sets are scalable (hundreds of kilobases to megabases), flexible and cost-effective; however, custom microarrays can obtain higher probe density (hundreds of thousands to millions of probes) to enable the capture of larger target regions [77]. Anecdotal evidence suggests that these methods differ considerably in efficiency, yet there remains tremendous room for optimization [78,79].

Current limitations of enrichment include biased database representation, leading to an inability to capture distantly related, extinct, divergent species/strains, novel genomic material, indels, plasmids, or exceptionally divergent regions. Ancestrally designed genomes (based on a phylogenetic tree) can be used to capture more distantly related taxa, those that are perceived to have diverged over longer time periods, or those that evolve more rapidly (RNA viruses).

In the future, advances in the recovery of nucleic acids from ancient remains and in the detection and capture of ancient pathogen DNA will involve reduction in costs and in the amount of material required. Such advances will also lead to improvements on success rates, which are currently below 0.01%.

Ancient Pathogen Presence and Virulence

Establishing the Presence of Specific Pathogens in Historical Populations

Disease symptoms, especially when inferred from historic records or skeletal remains are rarely diagnostic and have long generated doubts and debates about the nature of past disease and epidemic experience (Table 1). Demonstrating that a given

Box 3. Authenticating ancient pathogen DNA.

A range of methods can be employed to authenticate ancient pathogen sequences. These fall into eight categories:

1. ***DNA damage*.** Ancient DNA molecules are expected to be shorter (medians of 40–60 base pairs) than those from recent contaminating sources, although the distributions overlap. Current ancient DNA extractions and library preparations select for shorter and less damaged DNA molecules and thus the pool includes contaminating DNAs. Deamination patterns marked as C-to-T and G-to-A changes appear on the ends of authentic DNA molecules.
2. ***Random genome coverage*.** Mapped reads should be randomly and equally distributed across the genome (enriched at GC regions). Stacked, mapped reads are a tell-tale sign of multiple contributing sources of DNA.
3. ***Human DNA as benchmark*.** As most pathogen DNA is being reconstructed from the remains of human victims, using the presence and damage patterns of human DNA as a surrogate for their preservation is not without merit.
4. ***Mismatch distributions*.** Simple mismatch distributions, percent identities, or edit distances can be used to assess the proximity of the ancient reads to closely related contemporary genomes. This metric is database reliant. The expectation is that the majority of reads (barring damage) should be zero to a few mismatches from contemporary strains.
5. ***Haploidy of mapped reads*.** If sequencing depth is extensive, one can assess the overall contribution of singular or multiple circulating strains or even related pathogens through a comparison of the small number of variant sites using a site-frequency spectrum.
6. ***Comparison to contamination*.** There are ways to estimate the amount of contaminating modern human mitochondrial and nuclear DNA in ancient libraries [139]. Species identifications based on single-nucleotide polymorphisms or *k*-mers can help to distinguish soil-dwelling contaminants from real signals that can be further mined for genomic information.
7. ***Surrogates of microbial presence*.** Recovery of pathogen-specific protein or lipid markers from skeletal remains can support the identification of non-DNA products of pathogen infections (e.g., mycolic acids, proteins), immunological detection of antigens and/or the identification of organic biomolecules such as lipids [140].
8. ***Phylogenetics*.** Phylogenetic analysis can determine the relationship between ancient genomes and the modern diversity of the pathogen and a molecular-clock analysis can potentially be used to validate the age of the ancient sample itself [9,42]. Ancient genomes from lineages that are presumed to be now extinct will tend to sit outside the diversity of modern samples (Figure 2) [10,19]. In contrast, ancient genomes that are nested within the current diversity should be carefully assessed as possible contaminants.

microbe is present at a particular time and place can help to include that pathogen as a possible driver of a past mortality event. Characterizing pathogens genetically can also identify epidemic-associated, but clinically unconfirmed, opportunistic infections [12]. A dramatic illustration of the new ability to identify ancient DNA in a mass burial site came with the isolation of *Y. pestis* in a cemetery precisely dated to the Black Death (between the 1340s and 1350s) [5], generating new questions about the origins, persistence and transmission of recurrent plagues over subsequent centuries [80]. Similarly, isolates from mass burial sites related to the Roman Justinianic plague cycle (~540–750) assisted the identification of the pathogen associated with these important historical outbreaks [6,80–82]. Finally, plague ancient DNA recovered from western European sites of the 16th through 18th centuries showed the persistence of the initial Black Death strain of *Y. pestis* over several centuries [7,83,84]. Hence, the recovery and firm identification of ancient *Y. pestis* helped focus debates about the causative agent of these outbreaks that had developed due to the inevitable ambiguities of historical accounts.

Other notable ancient DNA findings linked to major mortality events include isolation of *Salmonella enterica* in association with *cocoliztli*, an indigenous Aztec term used to distinguish catastrophic die-offs from 1545 to the 1570s from the initial epidemics attributed to smallpox or measles [25]. A reconstructed strain of *Vibrio cholerae*, classical subtype, provided a more expected ancient DNA recovery from the second global pandemic of cholera in 1849 [25]. Important insights into the evolution of recent viral pandemics have been provided by influenza A/H1N1 virus recovered from paraffin-imbedded lung tissue of soldiers who died in the 1918/1919 influenza pandemic [25] and from strains of HIV subtype D from African victims during the early stages (1950s–1960s) of what would become the global HIV/AIDS epidemic [39,41].

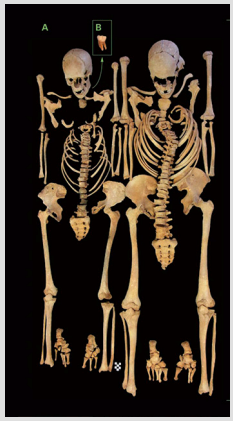
Evidence that links a given pathogen to a specific historical context is rare, particularly as multiple pathogens can co-circulate in populations. Indeed, attention to co-circulating pathogens in a given population or environment may provide a more complete understanding of past epidemics than does focussing on individual pathogenic causes [85]. Such an approach, known as 'pathocoenosis', has been applied to the study of ancient *Plasmodium* parasites [86,87].

Analysis of Virulence Determinants

Outbreaks of a specific disease can vary enormously in the extent of associated disease morbidity and mortality. Comparison of the genomes of microbes associated with very different case fatality rates provides a unique opportunity to explore whether these differences are due to genetically encoded virulence traits or environmental or host factors (or both), the presence or absence of key genetic features (e.g., antimicrobial resistance) and how virulence determinants evolve [88]. At a genomic level, virulence can be driven by the presence of plasmids, mobile genetic elements or single-nucleotide polymorphisms [88,89].

Increased access to ancient microbial genomes will help place constraints on the origin and timing of certain alleles in the population and enable, in some cases, the direct testing of their effect on pathogen adaptation [89]. Work on *Y. pestis* has shown

Table 1. How ancient pathogen DNA has revised (or not) historical or disease diagnoses.

Disease/epidemic/sample	Prior view of epidemic history	Did ancient DNA confirm view?
Black Death pandemic (1347–1353)  Six victims from the 1349–1350 East Smithfield Black Death cemetery (image with permission from Museum of London Archaeology).	Most believed *Y. pestis* caused the epidemics in Western Europe, although historical accounts provide conflicting evidence about dominant symptoms. The absence of *Y. pestis* in Western Europe today was an argument used to support further doubts.	*Y. pestis* draft genome reconstructed from human dental pulp in individuals buried in a medieval London cemetery used exclusively for Black Death victims during 1349–1350 [5].
Justinianic plague cycle (~540–750)  6th century skeletal remains of an adult male and female from Altenerding, Germany, positive for *Yersinia pestis* from the Justinian Plague (image with permission from Oxford University Press [97]).	Eyewitness accounts of the initial pandemic offer insupportably high mortality claims; narrative evidence from subsequent outbreaks also ambivalent.	A now extinct lineage of *Y. pestis* was retrieved from two late ancient burial sites [6,80,81].
Tuberculosis: *M. tuberculosis*  Classic tuberculosis damage to the spine of an ancient Peruvian, ~1000 years ago (image reprinted with permission from Springer Nature © 2014 from [2]).	Zoonotic emergence in early hominins predated *Homo sapiens* migrations to the Western Hemisphere, 20–13 thousand years ago. *M. bovis* from bovids was considered the likely precursor of *M. tuberculosis.*	Ancient genomic evidence supports a surprisingly young age for *M. tuberculosis*, and shows that *M. bovis* emerged even later. Origins of *M. tuberculosis* remain uncertain.

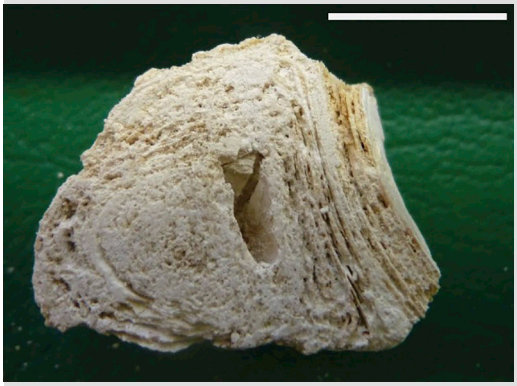
(*Continued on next page*)

Table 1. ***Continued*.**

Disease/epidemic/sample	Prior view of epidemic history	Did ancient DNA confirm view?
Smallpox  Sarcophagus with mummified child, mid-16th century Naples, Italy (image with permission from Gino Fornaciari from [9]).	Paleopathological diagnosis presumed smallpox via pox-like rash on skin and electron microscopic images of viral particles as well as immunoassays from a 16th century Italian mummy [141]. Contemporary with urban smallpox epidemics and catastrophic presumed smallpox in the Americas.	Discussion of early medical descriptions of smallpox epidemics in historical context, and the appearance of pocks on mummified human skin, were convincing evidence of VARV in the past. HBV, not VARV, found in a 16th century Italian mummy. The strain of HBV recovered belongs to genotype D, common in the Mediterranean region today [9].
Smallpox  Partial mummified remains of a child from Lithuanian crypt, late 17th century (image from [10]).	Emergence of VARV long assumed at least 3000 years ago (from electron microscopy of pock on the face of Pharaoh Ramesses V, ~1050 BCE). Historical peak in recorded global smallpox epidemics instead 1550–1750, suggesting a much later emergence of *Variola major.*	Virus recovered from a child mummy, c. 1643–1665, Vilnius, Lithuania [10], confirms most recent common ancestor of *V. major* emergence 1580–1630. Deeper relatives of smallpox, isolated from Early Middle Age including a Viking sample [103]. Animal pox virus that led to VARV emergence as well as region(s) and date of smallpox emergence thus far unknown.
Pathogen cause of catastrophic demographic collapse  Image from the Telleriano Remensis (Aztec) codex, depicting the great cocoliztli epidemic (image with permission from Springer Nature © 2018 from [18]).	The affliction called *cocoliztli* in Nahuatl is best translated as 'pestilence' and does not fully correspond closely to any modern clinical disease. Early Spanish and Nahuatl sources relevant to epidemics in 16th century Central Mexico diverge; *huey zahuatl* the more specific term for smallpox [142].	Rather than smallpox, evidence of *Salmonella enterica*, serovar Paratyphi C in 1545 site in Oaxaca, Mexico, lending support to the view that this mortality crisis was multicausal and associated with a megadrought [11].

(*Continued on next page*)

Table 1. ***Continued*.**

Disease/epidemic/sample	Prior view of epidemic history	Did ancient DNA confirm view?
Helicobacter pylori  Ötzi or the Iceman. The naturally mummified remains of a male from Alps ~3500 BCE (image: Thilo Part/ Wikimedia Commons (CC BY-SA 4.0)).	Decribed since the 1990s as following the global dispersal of *Homo sapiens* from Africa. Modern genetic diversity reflected human migration. No historical or paleopathological attention to this disease prior to the 1990s.	Metagenomic analysis of the ~5300-year-old Alpine Iceman's intestine showed an 'Asian' lineage that pre-dated the hybrid Asian/ African *H. pylori* ancestor of today's *hpEurope* lineage [13].
Mycobacterium leprae  Distinctive "bear claw" appearance of leprosy-damaged feet. Odense, Denmark skeletal remains (image used with permission from Science Museum, London).	Emergence of Hansen's disease long debated among paleopathologists. Historical debates center on causes for the sharp decline in leprosy cases, 1500s–1700s, Western Europe.	Genomic analysis of medieval leprosy cases (UK, Sweden, Denmark) showed very little change to the genome over the last 1000 years [1]. Debate about leprosy disappearance not linked to genome, thus ancient DNA findings support those who argue public health, refined medical diagnosis and hospital care have reduced prevalence of leprosy.
Unknown opportunistic pathogens  Calcified nodule, 12th century Troy (image: Caitlin Pepperell).	Much perinatal maternal mortality in the past assumed to be due to acute infections with common skin pathogens, *Streptococcus* spp. and *Staphylococcus* spp.	Both *S. saprophyticus* and *G. vaginalis* found in an abcess resulting from ectopic calcification, 13th century Troy. Opportunistic pathogens in the past enhanced overall mortality and contributed to differential mortality experience documented in many historical societies [12].

that it differs from *Yersinia pseudotuberculosis*, its closest modern relative, by deletions to the genome and the acquisition of two plasmids (pMT1 and pPCP1), which enabled its transmission from arthropods to mammalian hosts as bubonic plague. The finding that a 2,900-year-old *Y. pestis* genome already had these adaptations allows us some resolution on the origin and timing of the virulence mechanisms that led to high human fatality rates [11,90]. In contrast, ancient virulence factors found in a

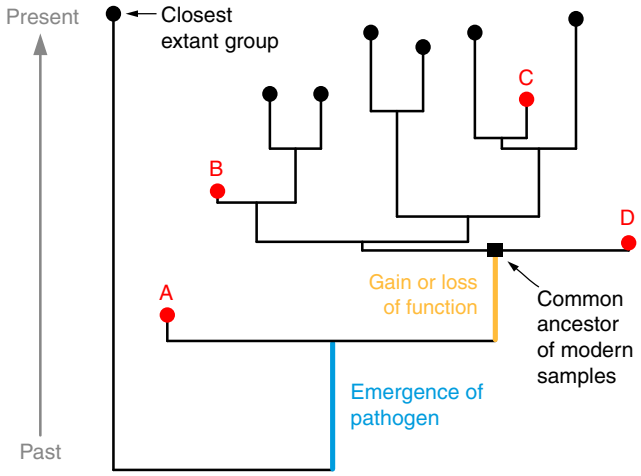


Figure 2. Phylogenetic analysis of ancient genomes.
Schematic phylogenetic tree showing how ancient genomes can provide information on key various aspects of the evolutionary history of pathogens. An ancient pathogen genome can potentially be placed as (A) an extinct sister lineage to the modern diversity (e.g., ancient variola virus from 7th–10th century [103]; image: teeth from an East Smithfield individual used to indicate source from which ancient variola-like sequences have been isolated, courtesy of Sharon DeWitte, Museum of London); (B) an extinct sister lineage to a modern haplogroup or genotype, but still falling within the modern clade of the pathogen (e.g., variola virus from the 17th century; image: VD21 child mummy from Vilnius, Lithuania, from which smallpox was detected, from [10]); (C) belonging to a present-day haplogroup or genotype (e.g., hepatitis B virus from the 16th century [9,125]; image: child mummy from Naples, Italy, with HBV detected, © 2018 Patterson Ross *et al.* CC BY 4.0); or (D) at the base of the modern clade, possibly as a direct ancestor (e.g., *Y. pestis* from 1348 [5]; image: skull of an individual from East Smithfield that yielded a Black Death genome, courtesy of Jelena Bekvalac, Museum of London (MIN86)). The square symbol denotes the common ancestor of modern pathogen samples. Importantly, the emergence of a pathogen in its present host could have occurred *at any point along* the branch between the divergence from its closest extant or extinct relative and the most recent common ancestor of the sampled isolates (branch in blue). Ancient pathogen genomes can help narrow this window of emergence [90] while also shedding light on any gain or loss of function along the branches leading up to the modern clade (branch in orange).

12th century genome of *S. enterica* Paratyphi C contained the same pathogenicity islands in mobile genetic elements as do modern strains, suggesting that the bacterium has caused similar disease over its 3,000-year history [19]. With both *S. enterica* and *Y. pestis*, the ancient presence of virulence factors indicates that subsequent decline in their incidence and disease severity was most likely driven by changes in human behaviour and pathogen ecology, rather than a decrease in their ability to spread and cause disease.

Evolutionary Rates and Timescales of Pathogens

One of the main goals of genomic and historical analyses of pathogens is to reconstruct the timescale of their emergence and spread. Human pathogens commonly emerge by jumping across a species barrier from animals (zoonoses) [91], although most of these jumps result in dead-end emergences and extinctions as was the case with the strain of *Y. pestis* responsible for the Plague of Justinian. A different strain of *Y. pestis* affected all areas of western Eurasia and north Africa during the second pandemic of plague, between the 14th and 19th centuries from regionally diverse rodent host foci. Occasionally, however, newly emerged pathogens establish sustained transmission cycles. Estimating the timing of the initial host switch, while interesting from a historical perspective, is challenging because this event could have happened at any point in the window of time between the divergence of the pathogen from its sister lineage and the most recent common ancestor of the sampled genomes of the pathogen (Figure 2) [92]. The bounds of this window can be estimated using molecular-clock analyses of genomic sequence data but can be very uncertain, which limits specific historical contextualization.

Reconstructing the evolutionary timescale of a pathogen relies on accurate estimation of the rate of its evolution. Pathogen genomes can undergo measurable amounts of evolutionary change between sampling times, which can span a period of weeks, years or even millennia [93,94]. Adding ancient genomes to data sets widens the sampling period, allowing a greater amount of evolutionary change to be evaluated (Figure 3) [95,96]. The assumption behind this approach is that evolutionary change occurs at a predictable rate over time, known as the 'molecular clock'. Unfortunately, pathogens with complicated epidemiology might evolve at highly variable rates and thus need careful evaluation before their genomes are used in evolutionary analyses, if molecular dating is to be attempted at all.

An important step in the phylogenomic analysis of pathogens is testing whether the genomic data carry a temporal signal. The simplest way to do this is to estimate a phylogenetic tree with branch lengths that are proportional to the observed number of substitutions along that branch and then to conduct a regression of the distance from the root to each of the tips as a function of the sampling times [97]. A positive slope suggests that the data set has a temporal signal (Figure 3), given that we expect samples with ancient DNA to have a smaller distance from the root than those with modern DNA (variola virus genomes show this nicely, for example). More sophisticated methods for evaluating the temporal signal and for estimating evolutionary rates have been developed over the past two decades [98]. The flexibility of these phylogenetic methods also allows the dating of ancient DNA samples for which the true age might be unknown [99], or to validate estimates of sample ages obtained using independent techniques [9,27].

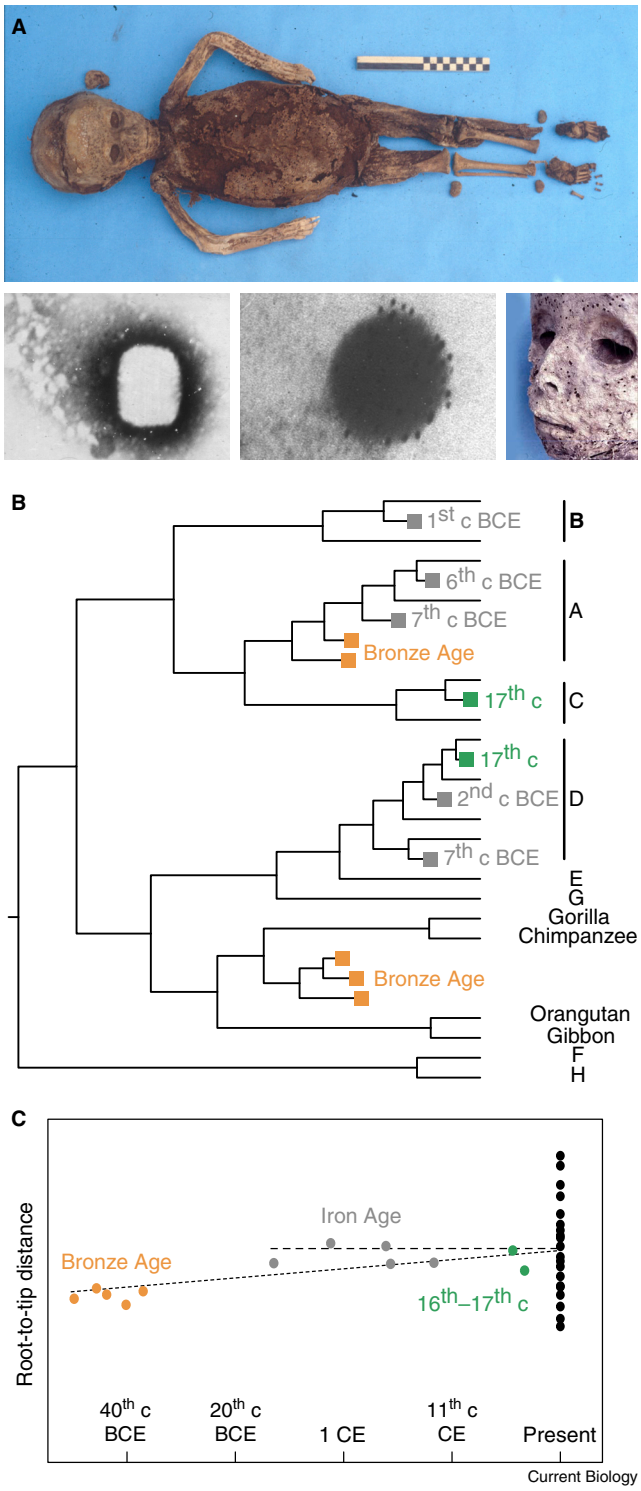


Figure 3. Conflict between the 16th century evidence for smallpox in Europe and the molecular evidence which finds HBV.
(A) Child mummy from the Basilica Saint Domenico Maggiore, Naples, Italy (1569 ± 60 years), prior to autopsy, vesicopustular rash on facial features and electron microscopy showing large, box-like viral particles as well as gold immune-stained (smallpox) particles from tissues (images with permission from Gino Fornaciari). (B) Phylogenetic tree showing placement of ancient reconstructed HBV strains from 16th century CE as well as Iron Age and Bronze Age samples, indicating that HBV has had a close association with humans for thousands of years. Letters indicate HBV assigned haplogroups. (C) Regression of root-to-tip genetic distances against sampling times, showing the absence of temporal signal in modern sequences and even with samples collected 2000 years before present (Iron Age), with a flat regression (slope = 0). Although the inclusion of Bronze Age sequences leads to a positive regression slope, the root-to-tip distances of these sequences are within the range of those of modern samples, such that extent of temporal signal remains inconclusive.

The main determinants of evolutionary rates are the type of nucleic acid (i.e., RNA or DNA), the genome size and the host. Viruses with single-stranded RNA genomes usually evolve orders of magnitude more rapidly than those with double-stranded DNA genomes [100]. The upshot of such evolutionary rate variation in microbial genomes is that the sampling period should be carefully considered according to the pathogen being studied. Sequence data collected from rapidly evolving RNA viruses over a few years are usually sufficiently informative to allow their evolutionary rate to be estimated. In contrast, for more slowly evolving bacteria such as *M. tuberculosis*, an ancient genome is sometimes essential to obtain sufficient temporal signal to estimate an evolutionary rate and timescale [101,102].

The divergence between a pathogen and its closest known relative places a maximum bound on the timing of its emergence. For example, sequences of *Variola virus* recovered from human remains from the Early Middle Ages (793–1066) fall outside the diversity of the lineages associated with modern smallpox and are separated from them by a long branch [103]. This phylogenetic pattern indicates that the initial cross-species transmission to humans must have occurred between the earliest detection of this virus — dating back to at least the Viking era — and its closest known animal relatives, camelpox from camels and taterapox from African gerbils [10] (Figure 2). Importantly, the 'Viking' age and modern smallpox Variola strains differ in the inactivated genes they possess. This feature suggests that smallpox, as we understand it epidemiologically today, is likely to have evolved at some point after the divergence of the Viking age Variola strains and the common ancestor of modern smallpox. Unfortunately, estimating this date can be seriously misled by the effects of purifying selection and misspecification of the evolutionary model. In particular, the removal of deleterious genetic variants by purifying selection can produce a time-dependent bias in evolutionary rate estimates, leading to underestimation of deep divergence times [102,104]. Adding ancient genomes to the data set widens the sampling period, potentially allowing for better inference of the evolutionary rate of the pathogen. However, if the ancient genomes represent divergent lineages that differ in their biological properties, the inclusion of these genomes can also introduce considerable heterogeneity that may require careful consideration of the molecular-clock model and abandoning hope of obtaining meaningful date estimates [105,106].

Ancient genomes are often the only way to uncover the existence of now-extinct pathogen lineages. For some pathogens, the most recent common ancestor of modern lineages is surprisingly young compared with perceptions of the age of the disease. This is especially true for pathogens that have experienced substantial reductions in genetic diversity, such as viruses that have been targeted by extensive vaccination programs [10,42]. Despite these challenges, molecular-clock

analyses can still be useful in placing bounds on when pathogens emerged and diversified. In this way, molecular dating provides a valuable means of testing hypotheses about pathogen origins and evolution.

Pathogen Phylogeography and Phylodynamics

The ancestral movements and present-day distributions of pathogens can be inferred by phylogeographic analysis, which combines geographic and evolutionary information. Simple phylogeographic methods overlay geographic information on the tips of phylogenetic trees to interpret how migration events could have given rise to the observed distribution. Formal statistical approaches are used to model past movements between geographic locations, which can be specified as a discrete (e.g., country or city) or a continuous trait (e.g., GPS coordinates) [107]. Importantly, phylogeographic models are sensitive to sampling bias; accurate inference of ancestral locations typically requires the inclusion of samples from the ancestral location itself [108], which is certainly unknown in the ancient context described herein. One of the benefits of ancient DNA is to provide evidence for or against the presumed ancestral origin of a particular strain or pathogen (Box 1). In some cases, ancient pathogen phylogeography has been used as a proxy for ancient human dispersal [109,110], sometimes providing results contrary to expectations defined by archaeological evidence alone [111].

The spatial distribution and emergence of pathogens that spread rapidly in host populations tend to have complex geographic histories that would be impossible to untangle without ancient DNA. How *Y. pestis* spread through rodent populations and caused repeated human spill-over epidemics in Europe and the Middle East between the 1340s and the 1840s has received substantial attention. As more than 100 ancient genomes of *Y. pestis* genomes have now been sequenced, it may be possible to inform historical debates through analyses of ancient DNA. For example, did European plague epidemics originate from repeated re-emergences from local rodent reservoirs [112] or did successive transmission waves reach Europe, initiated by great gerbil epizootics in the central Eurasian grasslands [113]? The discovery of *Y. pestis* genomes in mass burial sites in 18th century France revealed a lineage that is not present in the modern diversity of this bacterium. Instead, the lineage is linked to the 14th century one from England and thus supports the hypothesis of a presently extinct Mediterranean reservoir rather than the traditional narrative, that plague always 'came from the east' [84,85].

Many pathogens evolve at timescales similar to those over which they accumulate genetic changes, and their phylogenetic patterns provide meaningful insights into their epidemiological processes. These approaches have been combined in the field of phylodynamics [114], where various branching models have been proposed to infer epidemiological parameters. Most of their models are based on the coalescent, which models the ancestry of genes sampled from a population, or the birth–death process, which models speciation and extinction through time [115]. The inclusion of ancient DNA data brings an opportunity to infer the tempo and mode of previous pandemics, but it requires careful consideration of the model in question and its assumptions.

Coalescent and birth–death models with constant dynamics perform well in analyses of recently emerging outbreaks, but they are not realistic for the complex epidemiological dynamics of infectious diseases with long histories. For example, the persistent re-emergence of *Y. pestis* would not conform to such constant dynamics. A more flexible alternative is offered by skyline methods that can allow the epidemic growth rate, population size and other parameters to vary across different time intervals [116]. Recent developments in skyline techniques that can explicitly model the sampling process [117] and those that allow the inclusion of occurrence data with low-quality or no sequence information [118] offer promising avenues for coherent phylodynamic analyses of ancient DNA. Further research into the performance of these methods is needed to assess their accuracy in reconstructing past epidemics.

Ethics and the Hunt for Ancient Pathogen DNA

Despite the insights provided by ancient pathogen DNA, it is important to ask what the inherent value of an ancient pathogen genome is when it comes at the cost of partial destruction of a skeletal feature. As of this writing, 236 pathogen genomes have been reconstructed from 12,733 individual remains sampled, a success rate of 0.01%. Does this represent a reasonable rate of success, thereby justifying the extensive sampling? If not, what does and how is this value to be determined? While this ethical quandary is nothing new for the field of ancient DNA, nor archaeology more generally, it remains an important and unresolved issue. Curators rightly ask to see the results of a pilot project to determine the likelihood of overall success but measuring success can be elusive. Is it the isolation of human DNA from a certain percentage of human remains and, if so, what percentage is deemed a success in a pilot study and opens the way to broader sampling? Or do researchers need to find the presence of a notable pathogen prior to additional sampling? Neither of these guarantees the presence or retrievability of actual ancient pathogen genomes. Most ancient pathogen DNAs detected have come from scanning the 'left-over', non-human DNA reads of deeply sequenced libraries, which avoids additional destructive sampling but comes at a high financial cost.

With an increasing number of researchers in the field, competing for access to the same collections, the question of who gets to sample has increasing political ramifications. Should sampling permission be given to researchers who work (or propose to do so) on ancient human DNA, providing the raw sequencing data for others interested in pathogen research and *vice versa*, or only to research groups that are doing both? One thing that is clear is that before sampling can begin, stakeholders in any particular project or present-day descendants of the human remains being studied must agree on best practices, what and how much to sample and how best to integrate the DNA findings with existing written or oral histories. We also need to shed outdated historical narratives and the patterns of resource collection and storage that once drove Western colonial medicine and science. Fortunately, there is a strong precedent of how this can work, as illustrated by the SING consortium that trains indigenous scientists in genomics while bringing the indigenous perspective to the interpretation of the data [119]. However, unlike work for ancient human

DNA, most pathogen DNA work remains mysterious as pathogen discoveries have come almost entirely as surprise findings, making it difficult but presumably necessary to justify destructive sampling for hypothesis-driven projects.

Outlook

The recovery of ancient DNA has proven valuable for our understanding of how and when pathogens emerge and evolve. Early on, ancient DNA research relied on PCR and was used to query the dispersals of humans and other vertebrate species, adding some investigation of presumed tuberculosis or leprosy seen in skeletal remains. By the end of the 1990s, ancient DNA research had shifted to the phylogenetic relationships of particular pathogens prominent since the Neolithic [11,120]. The recent sequencing of ancient RNA viral genomes [42] is an exciting development that may revolutionize our understanding of the emergence and origin of many other RNA-based pathogens. Moreover, newer analytical tools to query huge data sets from the modern era can now provide a more comprehensive understanding of ancient pathogen diversity. These advances were unimaginable at the beginning of the current millennium but now challenge the field to consider how best to conserve some part of our collective past resources for future generations. In the future, to better understand and contextualize the dynamics that drove pathogen evolution, it will be essential to improve collaboration with scholars who study the remote to the recent past, including palaeoclimatologists, archaeologists and historians.

The past can never be fully retrieved, no matter how advanced our technologies. For example, despite the generation of unprecedented amounts of genome sequence data during the COVID-19 pandemic (~75,000 genomes in the GISAID database [121]) and a detailed understanding of the spread of its causative agent, the SARS-CoV-2 virus [122], the evolutionary history of the virus remains mysterious because of sparse sampling, multiple possible reservoirs and a complex history of recombination [123,124]. Rewind a few centuries and pathogen identification, origins and evolution become clouded even further. Fortunately, by accessing the genomes of ancient pathogens that were responsible for historical morbidity and mortality events, ancient DNA can provide a glimpse into the evolutionary and epidemiological past, amplifying our understanding of infectious disease emergence and thereby placing pandemics like COVID-19 into their broader context.

ACKNOWLEDGMENTS

H.P. acknowledges all current and former lab members for their thoughtful insights and tremendous input into the discovery of ancient molecules, without whom most of the work cited herein would never have come to fruition. H.P. also thanks D. Poinar for continued help on illuminating the evolutionary history of infectious disease, Ravneet Sidhu and Stephanie Marciniak for compiling the pathogen database and the numerous collaborators and curators who have graciously provided access to the many remains that have been used for study. H.P. is supported through the Social Sciences and Humanities, Research Council, Natural Sciences and Engineering Research Council, Canadian Institutes for Advanced Research, the Michael G. DeGroote Institute for Infectious Disease Research and McMaster University. S.D. was funded by the Australian Research Council (grant number: DE190100805) and the National Health and Medical Research Council (grant number: APP1157586).

REFERENCES

1. Schuenemann, V.J., Avanzi, C., Krause-Kyora, B., Seitz, A., Herbig, A., Inskip, S.A., Bonazzi, M., Reiter, E., Urban, C., Pedersen, D.D., *et al*. (2018). Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. PLoS Pathog. *14*, e1006997.

2. Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., *et al*. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature *514*, 494–497.

3. Marciniak, S., Prowse, T.L., Herring, D.A., Klunk, J., Kuch, M., Duggan, A.T., Bondioli, L., Holmes, E.C., and Poinar, H.N. (2016). *Plasmodium falciparum* malaria in 1st-2nd century CE southern Italy. Curr. Biol. *26*, R1220–R1222.

4. Green, M. (2017). The globalisations of disease. In Human Dispersal and Species Movement: From Prehistory to the Present, N. Boivin, R. Crassard, and M.D. Petraglia, eds. (Cambridge: Cambridge University Press), pp. 494–520.

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

6. Wagner, D.M., Klunk, J., Harbeck, M., Devault, A.M., Waglechner, N., Sahl, J.W., Enk, J.M., Birdsell, D.N., Kuch, M., Lumibao, C., *et al*. (2014). *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. Lancet Infect. Dis. *14*, 319–326.

7. Spyrou, M.A., Tukhbatova, R., Feldman, M., Drath, J., Beltrán de Heredia, J., Arnold, S., Sitdikov, A.G., Castex, D., Wahl, J., Gazimzyanov, I.R., *et al*. (2016). Historical Y. pestis genomes reveal the European Black Death as the source of ancient and modern plague pandemics. Cell Host Microbe *19*, 874–881.

8. Krause-Kyora, B., Nutsua, M., Boehme, L., Pierini, F., Pedersen, D.D., Kornell, S.-C., Drichel, D., Bonazzi, M., Möbus, L., Tarp, P., *et al*. (2018). Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nat. Commun. *9*, 1569.

9. Patterson Ross, Z., Klunk, J., Fornaciari, G., Giuffra, V., Duchêne, S., Duggan, A.T., Poinar, D., Douglas, M.W., Eden, J.-S., Holmes, E.C., *et al*. (2018). The paradox of HBV evolution as revealed from a 16th century mummy. PLoS Pathog. *14*, e1006750.

10. Duggan, A.T., Perdomo, M.F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Buchmann, J.P., Jankauskas, R., Humphreys, M., Golding, G.B., Southon, J., *et al*. (2016). 17th century variola virus reveals the recent history of smallpox. Curr. Biol. *26*, 3407–3412.

11. Key, F.M., Posth, C., Esquivel-Gomez, L.R., Hübler, R., Spyrou, M.A., Neumann, G.U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., *et al*. (2020). Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. Nat. Ecol. Evol. *4*, 324–333.

12. Devault, A.M., Mortimer, T.D., Kitchen, A., Kiesewetter, H., Enk, J.M., Golding, G.B., Southon, J., Kuch, M., Duggan, A.T., Aylward, W., *et al*. (2017). A molecular portrait of maternal sepsis from Byzantine Troy. eLife *6*, e20983.

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

14. Anastasiou, E., and Mitchell, P.D. (2013). Palaeopathology and genes: investigating the genetics of infectious diseases in excavated human skeletal remains and mummies from past populations. Gene *528*, 33–40.

15. Majander, K., Pfrengle, S., Neukamm, J., Kocher, A., du Plessis, L., Pla-Díaz, M., Arora, N., Akgül, G., Salo, K., Schats, R., *et al*. (2020). Ancient bacterial genomes reveal a formerly unknown diversity of *Treponema pallidum* strains in early modern Europe. bioRxiv, https://doi.org/10.1101/2020.06.09.142547.

16. DeWitte, S.N., and Hughes-Morey, G. (2012). Stature and frailty during the Black Death: the effect of stature on risks of epidemic mortality in London, A.D. 1348–1350. J. Archaeol. Sci. *39*, 1412–1419.

17. Giffin, K., Lankapalli, A.K., Sabin, S., Spyrou, M.A., Posth, C., Kozakaitė, J., Friedrich, R., Miliauskienė, Ž., Jankauskas, R., Herbig, A., *et al*. (2020). A treponemal genome from an historic plague victim supports a recent emergence of yaws and its presence in 15th century Europe. Sci. Rep. *10*, 9499.

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

19. Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M.J., Langridge, G., Fotakis, A.K., Nair, S., Stenøien, H.K., *et al*. (2018). 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.

20. Kay, G.L., Sergeant, M.J., Giuffra, V., Bandiera, P., Milanese, M., Bramanti, B., Bianucci, R., and Pallen, M.J. (2014). Recovery of a medieval *Brucella melitensis* genome using shotgun metagenomics. mBio *5*, e01337–14.

21. Austin, R.M., Sholts, S.B., Williams, L., Kistler, L., and Hofman, C.A. (2019). To curate the molecular past, museums need a carefully considered set of best practices. Proc. Natl Acad. Sci. USA *116*, 1471–1474.

22. Haldon, J., Mordechai, L., Newfield, T.P., Chase, A.F., Izdebski, A., Guzowski, P., Labuhn, I., and Roberts, N. (2018). History meets palaeoscience: consilience and collaboration in studying past societal responses to environmental change. Proc. Natl Acad. Sci. USA *115*, 3210–3218.

23. Eglinton, G., and Logan, G.A. (1991). Molecular preservation. Philos. Trans. R. Soc. Lond. B *333*, 315–328.

24. Pääbo, S., Poinar, H.N., Serrre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annu. Rev. Genet. *38*, 645–679.

25. Hansen, H.B., Damgaard, P.B., Margryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., and Allentoft, M.E. (2017). Comparing ancient DNA preservation in petrous bone and tooth cementum. PLoS One *12*, e0170940.

26. Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610–3618.

27. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

28. Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning and enzymatic amplification. Proc. Natl Acad. Sci. USA *86*, 1939–1943.

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

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

31. Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol *12*, 87.

32. Geigl, E.-M. (2002). On the circumstances surrounding the preservation and analysis of very old DNA. Archaeometry *44*, 337–342.

33. Burbano, H.A., Hodges, E., Green, R.E., Briggs, A.W., Krause, J., Meyer, M., Good, J.M., Maricic, T., Johnson, P.L.F., Xuan, Z., *et al*. (2010). Timing of human protein evolution as revealed by massively parallel capture of Neandertal nuclear DNA sequences. Science *328*, 723–725.

34. Carpenter, M.L., Buenrostro, J.D., Valdiosera Morales, C.E., Schroeder, H., Allentoft, M.E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., *et al*. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am. J. Hum. Genet. *95*, 852–864.

35. Devault, A.M., McLoughlin, K., Jaing, C., Gardner, S., Porter, T.M., Enk, J.M., Thissen, J., Allen, J., Borucki, M., DeWitte, S.N., *et al*. (2014). Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array. Sci. Rep. *4*, 4245.

36. Sirak, K., Fernandes, D., Cheronet, O., Harney, E., Mah, M., Mallick, S., Rohland, N., Adamski, N., Broomandkhoshbacht, N., Callan, K., *et al*. (2020). Human auditory ossicles as an alternative optimal source of ancient DNA. Genome Res. *30*, 427–436.

37. Ng, T.F.F., Chen, L.-F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P.D., Varsani, A., Kondov, N.O., Wong, W., Deng, X., *et al*. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. Proc. Natl Acad. Sci. USA *111*, 16842–16847.

38. Smith, O., Dunshea, G., Sindling, M.-H.S., Fedorov, S., Germonpre, M., Bocherens, H., and Gilbert, M.T.P. (2019). Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. PLoS Biol. *17*, e3000166.

39. Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M., and Ho, D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature *381*, 594–597.

40. Taubenberger, J.K., Reid, A.H., Krafft, A.E., Bijwaard, K.E., and Fanning, T.G. (1997). Initial genetic characterization of the 1918 "Spanish" influenza virus. Science *275*, 1793–1796.

41. Gryseels, S., Watts, T.D., Mpolesha, J.-M.K., Larsen, B.B., Lemey, P., Muyembe-Tamfum, J.-J., Teuwen, D.E., and Worobey, M. (2020). A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue. Proc. Natl Acad. Sci. USA *117*, 12222–12229.

42. Düx, A., Lequime, S., Patrono, L.V., Vrancken, B., Boral, S., Gogarten, J.F., Hilbig, A., Horst, D., Merkel, K., Prepoint, B., *et al*. (2020). Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science *368*, 1367–1370.

43. Setlow, P. (2016). Spore resistance properties. In The Bacterial Spore: From Molecules to Systems, A. Driks, and P. Eichenberger, eds. (Washington, DC: ASM Press), pp. 201–216.

44. Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., and Bos, K.I. (2013). Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science *341*, 179–183.

45. Cruz, A.R., Pillay, A., Zuluaga, A.V., Ramirez, L.G., Duque, J.E., Aristizabal, G.E., Fiel-Gan, M.D., Jaramillo, R., Trujillo, R., Valencia, C., *et al*. (2010). Secondary syphilis in Cali, Colombia: new concepts in disease pathogenesis. PLoS Negl. Trop. Dis. *4*, e690.

46. Kolman, C.J., Centurion-Lara, A., Lukehart, S., Owsley, D.W., and Tuross, N. (1999). Identification of *Treponema pallidum* subspecies *pallidum* in a 200-year-old skeletal specimen. J. Infect. Dis. *180*, 2060–2063.

47. von Hunnius, T.E., Yang, D., Eng, B., Waye, J.S., and Saunders, S.R. (2007). Digging deeper into the limits of ancient DNA research on syphilis. J. Archaeol. Sci. *34*, 2091–2100.

48. Schuenemann, V.J., Lankapalli, A.K., Barquera, R., Nelson, E.A., Hernández-Zaragoza, D.I., Acuña-Alonzo, V., Bos, K.I., Márquez-Morfín, L., Herbig, A., and Krause, J. (2018). Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains. PLoS Pathog. *12*, e0006447.

49. Burrell, C.J., Howard, C.R., and Murphy, F.A. (2017). Virion structure and composition. In Fenner and White's Medical Virology (Cambridge MA: Academic Press), pp. 27–37.

50. DeWitte, S.N., and Stojanowski, C.M. (2015). The osteological paradox 20 years later: past perspectives, future directions. J. Archaeol. Res. *2015*, 397–450.

51. Larsen, C.S. (2018). The bioarchaeology of health crisis: infectious disease in the past. Annu. Rev. Microbiol. *47*, 295–313.

52. Roberts, C.A. (2011). The bioarchaeology of leprosy and tuberculosis. In Social Bioarchaeology, S.C. Agarwal, and B.A. Glencross, eds. (Hoboken, NJ: John Wiley and Sons), pp. 252–281.

53. van Dorp, L., Gelabert, P., Rieux, A., de Manuel, M., de-Dios, T., Gopalakrishnan, S., Carøe, C., Sandoval-Velasco, M., Fregel, R., Olalde, I., *et al*. (2019). *Plasmodium vivax* malaria viewed through the lens of an eradicated European Strain. Mol. Biol. Evol. *37*, 773–785.

54. Duggan, A.T., Klunk, J., Porter, A.F., Dhody, A.N., Hicks, R., Smith, G.L., Humphreys, M., McCollum, A.M., Davidson, W.B., Wilkins, K., *et al*. (2020). The origins and genomic diversity of American Civil War era smallpox vaccine strains. Genome Biol. *21*, 175.

55. Spyrou, M.A., Bos, K.I., Herbig, A., and Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. Nat. Rev. Genet. *20*, 323–340.

56. Kolling, G., Wu, M., and Guerrant, R.L. (2012). Enteric pathogens through life stages. Front. Cell. Infect. Microbiol. *2*, 114.

57. Devault, A.M., Golding, G.B., Waglechner, N., Enk, J.M., Kuch, M., Tien, J.H., Shi, M., Fisman, D.N., Dhody, A.N., Forrest, S.A., *et al*. (2014). Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. N. Engl. J. Med. *370*, 334–340.

58. Eng, S.-K., Pusparajah, P., Ab Mutalib, N.-S., Ser, H.-L., Chan, K.-G., and Lee, L.-H. (2015). *Salmonella:* a review on pathogenesis, epidemiology and antibiotic resistance. Front. Life Sci. *8*, 284–293.

59. Price, L.B., Hungate, B.A., Koch, B.J., Davis, G.S., and Liu, C.M. (2017). Colonizing opportunistic pathogens (COPs): the beasts in all of us. PLoS Pathog *13*, e1006369.

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

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

62. Marciniak, S., and Poinar, H.N. (2018). Ancient pathogens through human history: a paleogenomic perspective. In Paleogenomics, C. Lindqvist, and O.P. Rajora, eds. (Springer International Publishing), pp. 115–138.

63. Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R.D.E., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A.F., *et al*. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science *311*, 392–394.

64. Knapp, M., and Hofreiter, M. (2010). Next generation sequencing of ancient DNA: requirements, strategies and perspectives. Genes *1*, 227–243.

65. Millar, C.D., Huynen, L., Subramanian, S., Mohandesan, E., and Lambert, D.M. (2008). New developments in ancient genomics. Trends Ecol. Evol. *23*, 386–393.

66. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., *et al*. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl Acad. Sci. USA *110*, 15758–15763.

67. Gansauge, M.-T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., and Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. Nucleic Acids Res. *45*, e79.

68. Barlow, A., Forges, G.G., Dalén, L., Pinhasi, R., Gasparyan, B., Rabeder, G., Frischauf, C., Paijmans, J.L.A., and Hofreiter, M. (2016). Massive influence of DNA isolation and library preparation approaches on palaeogenomic sequencing data. bioRxiv, https://doi.org/10.1101/075911.

69. Troll, C.J., Kapp, J., Rao, V., Harkins, K.M., Cole, C., Naughton, C., Morgan, J.M., Shapiro, B., and Green, R.E. (2019). A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. BMC Genom. *20*, 1023.

70. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. *17*, 377–386.

71. Hübler, R., Key, F.M., Warinner, C., Bos, K.I., Krause, J., and Herbig, A. (2019). HOPS: automated detection and authentication of pathogen DNA in archaeological remains. Genome Biol. *20*, 280.

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

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

74. Warinner, C., Herbig, A., Mann, A., Yates, J.A.F., Weiß, C.L., Burbano, H.A., Orlando, L., and Krause, J. (2017). A robust framework for microbial archaeology. Annu. Rev. Genom. Hum. Genet. *18*, 321–356.

75. Wylie, T.N., Wylie, K.M., Herter, B.N., and Storch, G.A. (2015). Enhanced virome sequencing using targeted sequence capture. Genome Res. *25*, 1910–1920.

76. Bos, K.I., Jäger, G., Schuenemann, V.J., Vågene, Å.J., Spyrou, M.A., Herbig, A., Nieselt, K., and Krause, J. (2015). Parallel detection of ancient pathogens via array-based DNA capture. Philos. Trans. R. Soc. Lond. B, Biol. Sci. *370*, 20130375.

77. Orlando, L., Gilbert, M.T.P., and Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. Nat. Rev. Genet. *16*, 395–408.

78. Ávila-Arcos, M.C., Sandoval-Velasco, M., Schroeder, H., Carpenter, M.L., Malaspinuas, A.-S., Wales, N., Peñaloza, F., Bustamante, C.D., and Gilbert, M.T. (2015). Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. Meth. Ecol. Evol. *6*, 725–734.

79. Cruz-Dávalos, D.I., Llamas, B., Guanitz, C., Fages, A., Gamba, C., Soubrier, J., Librado, P., Seguin-Orlando, A., Pruvost, M., Alfarhan, A.H., *et al*. (2017). Experimental conditions improving in-solution target enrichment for ancient DNA. Mol. Ecol. Resour. *17*, 508–522.

80. Feldman, M., Harbeck, M., Keller, M., Spyrou, M.A., Rott, A., Trautmann, B., Scholz, H.C., Päffgen, B., Peters, J., McCormick, M., *et al*. (2016). A high-coverage *Yersinia pestis* genome from a sixth-century Justinianic plague victim. Mol. Biol. Evol. *33*, 2911–2923.

81. Keller, M., Spyrou, M.A., Scheib, C.L., Neumann, G.U., Kröpelin, A., Haas-Gebhard, B., Päffgen, B., Haberstroh, J., Ribera i Lacomba, A., Raynaud, C., *et al*. (2019). Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). Proc. Natl Acad. Sci. USA *116*, 12363–12372.

82. Harbeck, M., Seifert, L., Hänsch, S., Wagner, D.M., Birdsell, D., Parise, K.L., Wiechmann, I., Grupe, G., Thomas, A., Keim, P., *et al*. (2013). *Yersinia pestis* DNA from skeletal remains from the 6th century AD reveals insights into Justinianic plague. PLoS Pathog. *9*, e1003349.

83. Spyrou, M.A., Keller, M., Tukhbatova, R., Scheib, C.L., Nelson, E.A., Valtueña, A.A., Neumann, G.U., Walker, D., Alterauge, A., Carty, N., *et al*. (2019). Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. Nat. Commun. *10*, 4470.

84. Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M., *et al*. (2016). Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. eLife *5*, e12994.

85. Dutour, O. (2013). Paleoparasitology and paleopathology. Synergies for reconstructing the past of human infectious diseases and their pathocenosis. Int. J. Paleopathol. *3*, 145–149.

86. Grmek, M.D. (1969). Préliminaire d'une étude historique des maladies. Annales: E.S.C. *24*, 1437–1483.

87. Faure, E. (2014). Malarial pathocoenosis: beneficial and deleterious interactions between malaria and other human diseases. Front. Physiol. *5*, 441.

88. Andam, C.P., Worby, C.J., Chang, Q., and Campana, M.G. (2016). Microbial genomics of ancient plagues and outbreaks. Trends Microbiol. *24*, 978–990.

89. Fukuto, H.S., Vadyvaloo, V., McPhee, J.B., Poinar, H.N., and Bliska, J.B. (2018). A single amino acid change in the response regulator PhoP, acquired during *Yersinia pestis* evolution, affects PhoP target gene transcription and Polymyxin B Susceptibility. J. Bacteriol. *200*, e00050–18.

90. Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A.G., Schubert, M., Van Dam, A., Outzen Kapel, C.M., *et al*. (2015). Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. Cell *163*, 571–582.

91. Taylor, L.H., Latham, S.M., and Woolhouse, M.E.J. (2001). Risk factors for human disease emergence. Philos. Trans. R. Soc. Lond. B *356*, 983–989.

92. Ho, S.Y.W., and Duchêne, S. (2020). Dating the emergence of human pathogens. Science *368*, 1310–1311.

93. Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., and Rodrigo, A.G. (2003). Measurably evolving populations. Trends Ecol. Evol. *18*, 481–488.

94. Biek, R., Pybus, O.G., Lloyd-Smith, J.O., and Didelot, X. (2015). Measurably evolving pathogens in the genomic era. Trends Ecol. Evol. *30*, 306–313.

95. Duchêne, S., Lemey, P., Stadler, T., Ho, S.Y.W., Duchêne, D., Dhanasekaran, V., and Baele, G. (2020). Bayesian evaluation of temporal signal in measurably evolving populations. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msaa163.

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

97. Buonagurio, D.A., Nakada, S., Parvin, J.D., Krystal, M., Palese, P., and Fitch, W.M. (1986). Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science *232*, 980–982.

98. Ho, S.Y.W., and Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. Mol. Ecol. *23*, 5947–5965.

99. Shapiro, B., Ho, S.Y.W., Drummond, A.J., Suchard, M.A., Pybus, O.G., and Rambaut, A. (2011). A Bayesian phylogenetic method to estimate unknown sequence ages. Mol. Biol. Evol. *28*, 879–887.

100. Duchêne, S., Holmes, E.C., and Ho, S.Y.W. (2014). Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc. R. Soc. B *281*, 20140732.

101. Menardo, F., Duchêne, S., Brites, D., and Gagneux, S. (2019). The molecular clock of *Mycobacterium tuberculosis*. PLoS Pathog. *15*, e1008067.

102. Sabin, S., Herbig, A., Vågene, Å.J., Ahlström, T., Bozovic, G., Arcini, C., Kühnert, D., and Bos, K.I. (2019). A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex. bioRxiv, https://doi.org/10.1101/588277.

103. Mühlemann, B., Vinner, L., Margaryan, A., Wilhemson, H., de la Fuente Castro, C., Allentoft, M.E., Damgaard, P.B., Hansen, A.J., Nielsen, S.H., Strand, L.M., *et al*. (2020). Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science *369*, eaaw8977.

104. Duchêne, S., Holt, K.E., Weill, F.-X., Le Hello, S., Hawkey, J., Edwards, D.J., Fourment, M., and Holmes, E.C. (2016). Genome-scale rates of evolutionary change in bacteria. Microb. Genom. *2*, e000094.

105. Aiewsakun, P., and Katzourakis, A. (2016). Time-dependent rate phenomenon in viruses. J. Virol. *90*, 7184–7195.

106. Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Zu, L., *et al*. (2013). Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. Proc. Natl Acad. Sci. USA *110*, 577–582.

107. Zhukova, A., Gascuel, O., Duchêne, S., Ayers, D., Lemey, P., and Baele, G. (2020). Efficiently analysing large viral data sets in computational phylogenomics. In Phylogenetics in the Genomic Era, C. Scornavacca, F. Delsuc, and N. Galtier, eds. (no commercial publisher), pp. 5.3:1–5.3:43.

108. Kalkauskas, A., Perron, U., Sun, Y., Goldman, N., Baele, G., Guindon, S., and De Maio, N. (2020). Sampling bias and model choice in continuous phylogeography: getting lost on a random walk. bioRxiv, https://doi.org/10.1101/2020.02.18.954057.

109. Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.-Y., Maady, A., Bernöft, S., Thiberge, J.-M., Phuanukoonnon, S., *et al*. (2009). The peopling of the Pacific from a bacterial perspective. Science *323*, 527–530.

110. Achtman, M. (2016). How old are bacterial pathogens? Proc. R. Soc. B *283*, 20160990.

111. Yuen, L.K.W., Littlejohn, M., Duchêne, S., Edwards, R., Bukulatjpi, S., Binks, P., Jackson, K., Davies, J., Davis, J.S., Tong, S.Y.C., *et al*. (2019). Tracing ancient human migrations into Sahul using Hepatitis B virus genomes. Mol. Biol. Evol. *36*, 942–954.

112. Carmichael, A.G. (2014). Plague persistence in Western Europe: a hypothesis. In Pandemic Disease in the Medieval World: Rethinking the Black Death, M. Green, ed. (Kalamazoo, MI, USA, Bradford, UK: Arc Humanities Press), pp. 157–192.

113. Schmid, B.V., Büntgen, U., Easterday, W.R., Ginzler, C., Walløe, L., Bramanti, B., and Stenseth, N.C. (2015). Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. Proc. Natl Acad. Sci. USA *112*, 3020–3025.

114. Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J.A., and Holmes, E.C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. Science *303*, 327–332.

115. du Plessis, L., and Stadler, T. (2015). Getting to the root of epidemic spread with phylodynamic analysis of genomic data. Trends Microbiol. *23*, 383–386.

116. Pybus, O.G., and Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. Nat. Rev. Genet. *10*, 540–550.

117. Parag, K.V., du Plessis, L., and Pybus, O.G. (2020). Jointly inferring the dynamics of population size and sampling intensity from molecular sequences. Mol. Biol. Evol. *37*, 2414–2429.

118. Duchêne, S., Di Giallonardo, F., Holmes, E.C., and Vaughan, T.G. (2019). Inferring infectious disease phylodynamics with notification data. bioRxiv, https://doi.org/10.1101/596700.

119. Bardill, J., Baer, A.C., Garrison, N.A., Bolnick, D.A., Raff, J.A., Walker, A., and Malhi, R.S. (2018). Advancing the ethics of paleogenomics. Science *360*, 384–385.

120. Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. (1999). *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc. Natl Acad. Sci. USA *96*, 14043–14048.

121. Elbe, S., and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob. Chall. *1*, 33–46.

122. Duchêne, S., Featherstone, L., Haritopoulou-Sinanidou, Rambaut, A., Lemey, P., and Baele, G. (2020b). Temporal signal and the phylodynamic threshold of SARS-CoV-2. bioRxiv, https://doi.org/10.1101/2020.05.04.077735.

123. Boni, M.F., Lemey, P., Jiang, X., Lam, T.T., Perry, B., Castoe, T., Rambaut, A., and Robertson, D.L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol., https://doi.org/10.1038/s41564-020-0771-4.

124. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nat. Med. *26*, 450–452.

125. Blumberg, B.S. (2002). Hepatitis B: The Hunt for a Killer Virus (Princeton University Press).

126. Kahila Bar-Gal, G., Kim, M.J., Klein, A., Shin, D.H., Oh, C.S., Kim, J.W., Kim, T.-H., Kim, S.B., Grant, P.R., Pappo, O., *et al*. (2012). Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology *56*, 1671–1680.

127. Paraskevis, D., Magiorkinis, G., Magiokinis, E., Ho, S.Y.W., Belshaw, R., Allain, J.-P., and Hatzakis. (2012). Dating the origin and dispersal of hepatitis B virus infection in humans and primates. Hepatology *57*, 908–916.

128. Mühlemann, B., Jones, T.C., Damgaard, P.B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., *et al*. (2018). Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature *557*, 418–423.

129. Krause-Kyora, B., Susat, J., Key, F.M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D., Franzenburg, S., *et al*. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. eLife *7*, e36666.

130. Barquera, R., Lamnidis, T.C., Lankapalli, A.K., Kocher, A., Hernández-Zaragoza, D.I., Nelson, E.A., Zamora-Herrera, A.C., Ramallo, P., Bernal-Felipe, N., Immel, A., *et al*. (2020). Origin and health status of first-generation Africans from early colonial Mexico. Curr. Biol. *30*, 2078–2091.e2011.

131. Zadoks, J.C. (2008). The potato murrain on the European continent and the revolutions of 1848. Potato Res *51*, 5–45.

132. Mokyr, J., and Ó Gráda, C. (2002). What do people die of during famines: the Great Irish Famine in comparative perspective. Eur. Rev. Econ. Hist. *6*, 339–363.

133. Martin, M.D., Cappellini, E., Samaniego, J.A., Zepeda, M.L., Campos, P.F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S.Y.W., Dietrich, F.S., *et al*. (2013). Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. Nat. Commun. *4*, 2172.

134. Martin, M.D., Ho, S.Y.W., Wales, N., Ristaino, J.B., and Gilbert, M.T.P. (2014). Persistence of the mitochondrial lineage responsible for the Irish Potato Famine in extant New World *Phytophthora infestans*. Mol. Biol. Evol. *31*, 1414–1420.

135. Martin, M.D., Viera, F.G., Ho, S.Y.W., Wales, N., Schubert, M., Seguin-Orlando, A., Ristaino, J.B., and Gilbert, M.T. (2016). Genomic characterization of a South American *Phytophthora* hybrid mandates reassessment of the geographic origins of *Phytophthora infestans*. Mol. Biol. Evol. *33*, 478–491.

136. Yoshida, K., Schuenemann, V.J., Cano, L.M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F.N., Kamoun, S., Krause, J., *et al*. (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. eLife *2*, e00731.

137. Goss, E.M., Tabima, J.F., Cooke, D.E.L., Restrepo, S., Fry, W.E., Forbes, G.A., Fieland, V.J., Cardenas, M., and Grünwald, N.J. (2014). The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. Proc. Natl Acad. Sci. USA *111*, 8791–8796.

138. Rytting, J.R. (2000). A Disputacioun Betwyx þe Body and Wormes: a translation. Comitatus *31*, 217–232.

139. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

140. Tran, T.-N.-N., Aboudharam, G., Raoult, D., and Drancourt, M. (2011). Beyond ancient microbial DNA: nonnucleotidic biomolecules for paleomicrobiology. BioTechniques *50*, 370–380.

141. Fornaciari, G., and Marchetti, A. (1986). Intact smallpox virus particles in an Italian mummy of sixteenth century. Lancet *2*, 8521–8522.

142. McCaa, R. (1995). Spanish and Nahuatl views on smallpox and demographic catastrophe in Mexico. J. Interdiscip. Hist. *25*, 397–431.