Trends in Microbiology

Microbial Genomics of Ancient Plagues and Outbreaks

--Manuscript Draft--

Manuscript Number:	TIMI-D-16-00114R1
Article Type:	Review
Corresponding Author:	Cheryl P Andam, Ph.D Harvard T. H. Chan School of Public Health Boston, MA UNITED STATES
First Author:	Cheryl P Andam, Ph.D
Order of Authors:	Cheryl P Andam, Ph.D
	Colin J Worby, Ph.D.
	Qiuzhi Chang
	Michael G Campana, Ph.D.
Abstract:	The recent use of next generation sequencing methods to investigate historical disease outbreaks has provided us with an unprecedented ability to address important and long-standing questions in epidemiology, pathogen evolution and human history. In this review, we present major findings that illustrate how microbial genomics has provided new insights into the nature and etiology of infectious diseases of historical importance, such as plague, tuberculosis, and leprosy. Sequenced isolates collected from archaeological remains also provide evidence for the timing of historical evolutionary events as well as geographic spread of these pathogens. Elucidating the genomic basis of virulence in historical diseases can provide relevant information on how we can effectively understand the emergence and re-emergence of infectious diseases today and in the future.

Trends

- Important challenges to ancient genomic analyses include limited DNA sampling and methodological issues (DNA authentication, recovery, isolation, enrichment, sequencing, false positives).
- Genome sequencing of pathogens from historically notable disease outbreaks provides insight into the nature of long-term co-evolution of humans and pathogens.
- Microbial genomics can inform us of the origins and geographic spread of ancient diseases, which can be used to corroborate historical literature and archaeological evidence.
- Genomic data also provide information on the underlying causes of virulence and transmission of ancient pathogens, as well as relationship between historical and contemporary lineages.

Microbial Genomics of Ancient Plagues and Outbreaks

Cheryl P. Andam[1], Colin J. Worby[1], Qiuzhi Chang[1], Michael G. Campana[2]

[1] Harvard T. H. Chan School of Public Health, Department of Epidemiology, Boston, MA 02115

[2] Smithsonian Conservation Biology Institute, Center for Conservation Genomics, 3001 Connecticut Avenue NW, Washington, DC 20008

Correspondence:

candam@hsph.harvard.edu (C. P. Andam)

campanam@si.edu (M. G. Campana)

Keywords:

Genomics, genome, pathogen, plague, epidemic, ancient DNA

Abstract:

The recent use of next generation sequencing methods to investigate historical disease outbreaks has provided us with an unprecedented ability to address important and long-standing questions in epidemiology, pathogen evolution and human history. In this review, we present major findings that illustrate how microbial genomics has provided new insights into the nature and etiology of infectious diseases of historical importance, such as plague, tuberculosis, and leprosy. Sequenced isolates collected from archaeological remains also provide evidence for the timing of historical evolutionary events as well as geographic spread of these pathogens. Elucidating the genomic basis of virulence in historical diseases can provide relevant information on how we can effectively understand the emergence and re-emergence of infectious diseases today and in the future.

Genomics of Ancient Diseases: An Overview

Disease outbreaks and epidemics have always been part of human history and civilization. At different periods of our history, communicable diseases have decimated large proportions of the human population, greatly altering the structure and dynamics of societies and devastating economies (Figure 1) [1,2]. Centuries ago, the spread of disease was believed to be due to a noxious property of the air, or miasma, originating from rotting organic matter. Only after major discoveries in microbiology, such as Louis Pasteur's work on puerperal fever and *Vibrio* and Robert Koch's work on anthrax in the 19th century, was the miasma theory finally discarded and microorganisms were finally recognized as agents of diseases [3,4]. Remarkable scientific advances in microbiology have greatly enhanced our ability to detect and identify pathogens, and in turn, implement more effective treatments and infection control.

More recently, pathogen genome sequencing has been used as an epidemiological tool to trace contemporary outbreaks, as demonstrated during the cholera outbreak in Haiti in 2010 [5–7] and the Ebola epidemic in West Africa in 2014 [8–10]. Unsurprisingly, the vast majority of pathogen sequencing efforts have been focused on contemporary cases, due in large part to the increased difficulty of collecting and successfully sequencing historical pathogen genetic material, which degrades considerably over time. However, it has recently been demonstrated that this is indeed possible, and a number of studies have collected pathogen sequence data from ancient sources (Table 1). Previous studies of ancient epidemics have largely relied on historical and archaeological data alone, such as ancient texts, engraving, images and other relics as well as church records, wills and burial registers.

Such sources may often be inconclusive or ambiguous, and using them to support epidemiological hypotheses can be misleading. Genomic data can provide an additional source of information which can be considered alongside other evidence. For instance, the epidemiological dynamics of the medieval Black Death based on such records were consistent with a viral pathogen, spreading via aerosol or direct contact between persons [11]. Genetic analyses later confirmed the bacterium *Yersinia pestis* as the etiological agent. With the application of genome sequencing, we are now able to contribute new perspectives to many long-held questions on the etiology of past infectious diseases and gain insights into the nature of human-pathogen interactions through time (Figure 2).

In this review, we highlight the logistical and analytical challenges in recovering and sequencing ancient DNA, before describing some of the major findings that illustrate how microbial genomics has revolutionized our ability to investigate the nature of infectious diseases of historical importance. We focus on *Y. pestis*, being the pathogen which has been most frequently collected, sequenced and studied in a historical context. We also consider those diseases that have caused a significant burden in terms of morbidity and mortality from antiquity to early 20th century, namely plague, tuberculosis, leprosy, influenza and cholera. We highlight the contributions of genomics to our understanding of the origins and emergence of historical outbreaks, transmission routes, and the genetic diversity of ancient pathogens and their relationship to contemporary strains (Box 1). We conclude with how microbial genomics can be used as an important component in understanding the etiology of ancient diseases, and findings should be viewed alongside archaeological and historical evidence.

Challenges and Limitations to Ancient Pathogen Genomics

There have been huge advances in next generation sequencing technology in the past 15-20 years, and collecting genomic data from contemporary pathogen samples is now relatively quick, inexpensive and easy. While such advances have also facilitated the potential for sequencing ancient genomes, a number of additional challenges restrict the scale and feasibility of such data collection. While refinements in DNA sampling, recovery and sequencing methods continue to be improved upon (Box 1), it is important to consider sources of error and uncertainty when drawing conclusions.

With the notable exceptions of the 1918 Spanish flu virus [12], hepatitis B virus (HBV) [13] and HIV [14], almost all ancient pathogen genome studies are limited to DNA pathogens, since RNA is not typically expected to survive in the archaeological record [15]. Ancient DNA suffers from oxidative and hydrolytic damage and fragmentation into pieces typically shorter than 100 base pairs [16,17]. The most frequently observed damage artifact is deamination of cytosines to uracils, which tend to accumulate on the 3′ ends of ancient DNA molecules [18]. Analysis of these characteristic ancient DNA damage patterns has become commonplace to ensure reliable results since contaminating molecules are unlikely to show these signatures [19]. Additionally, upon death, tissues are colonized by fungal and bacterial decomposers, thus greatly reducing the relative quantity of endogenous molecules and may contribute to false positives. Ancient DNA yields for the host (let alone target pathogens) are

116 often less than 1% of the total DNA content (e.g., [20]). These factors complicate DNA
117 extraction, high-throughput sequencing library construction, DNA alignment and genome
118 analysis. Novel protocols, such as improved silica-based ancient DNA extraction [16],
119 single-stranded library construction [21], targeted nucleic acid capture (e.g. [22]) and partial
120 uracil-DNA-glycosylase digestion [19], have been developed to maximize high-quality
121 endogenous DNA yields. Ancient-DNA-specific pipelines, such as PALEOMIX [23] and
122 EAGER [24], have also been designed to generate and analyze ancient genomic data.
123
124 Most pathogens do not leave diagnostic changes in bone and teeth, the most commonly
125 preserved archaeological tissues [25]. Therefore, most studies have used material securely
126 linked by historical records to known pathogens (e.g., [22,26]). Dental calculus is also a rich
127 source for pathogens and the oral microbiome [27,28]. Identification of pathogens in
128 archaeological specimens without *a priori* diagnoses is particularly difficult since the small
129 quantities of target pathogen are difficult to distinguish from the contaminant background
130 [29]. Analysis of appropriate negative controls (such as soil or uninfected individuals from
131 similar archaeological contexts) is required to weed out false positives due to ubiquitous soil
132 and laboratory contaminants, which are often closely related to pathogens of interest [29].
133 DNA capture using diagnostic pathogen arrays has been proposed to identify unknown
134 pathogens [30], although the application of this technology remains limited. Pathogen
135 capture approaches produce higher yields of pathogen nucleic acids compared to shotgun
136 sequencing approaches (thereby improving genome sequencing coverage and experimental
137 cost-effectiveness). However, capture assays are biased towards known variation and extant
138 diversity (e.g., [31]). Additionally, short ancient DNA sequences complicate genome

assembly and analysis of evolutionary events such as genome rearrangements and gene duplications [32]. Furthermore, the possibility of mixed infections [33,34] compounds these analytical challenges.

Sampling remains a major limitation to making phylogenetic and transmission inferences, and this is particularly critical in ancient pathogens whereby sample collection is greatly hindered by the discovery and quality of archaeological remains. While phylogenetic methods provide a powerful framework to investigate evolutionary history, non-random sampling can disrupt inference of divergence times and tree structure considerably. Due to the numerous challenges associated with identifying and collecting ancient DNA, it is unlikely that sufficient samples can be collected to adequately represent the pathogen genetic diversity during historical outbreaks. Existing studies have necessarily been restricted to small numbers of often closely related samples (Table 1). While this does not impact many of the conclusions that may be drawn from examining the phylogenetic relationship between ancient and modern strains (i.e., whether extant strains are directly descended from historical lineages or the relative genomic changes between different outbreaks), the ascertainment of divergence times and the relationship between more closely related outbreaks (spatially and/or temporally) remains highly uncertain without a larger and representative sample.

Ancient Origins and Geographic Spread: the Case of the *Yersinia* Plague

Overview of the three Yersinia Plague Pandemics

As the most widely collected and sequenced ancient pathogen, *Y. pestis*, the causative agent of plague, illustrates most clearly the potential for ancient pathogen genomics to shed light on historical epidemiological and evolutionary dynamics. In this section, we describe some of the studies which have utilized genomic data to test hypotheses about the devastating spread of *Y. pestis* worldwide.

Y. pestis is a Gram-negative bacterium that can be transmitted from rodents to humans. In humans, it can cause bubonic, septicemic and pneumonic diseases, which are highly lethal without treatment with antibiotics. From historical records, it is known that at least three *Y. pestis* pandemics swept through large regions of the globe in the past 1,500 years. The first pandemic, also known as the Plague of Justinian, first occurred in Egypt and spread north to Constantinople in 541-543, the eastern capital of the Roman Empire, and neighboring regions. Across Europe and the Mediterranean, local outbreaks and dissemination occurred until the 8th century [35–37]. The second pandemic, widely known as the Black Death or the Great Pestilence, occurred in 1346-1352 and reduced the population of Europe by as much as 30 million, with resurgences of plague until the 18th century [38,39]. The third pandemic is believed to have begun in southwestern China in the 1860s, and its spread to other countries was accelerated through the use of steamships, railways and the movement of military troops [40,41]. While archaeological remains recovered over many regions in Europe and Asia have allowed researchers to make significant strides in understanding the human and social impacts of *Y. pestis*, the availability of genomic data has significantly enhanced our knowledge of its origins and evolutionary history.

Phylogenetic Relationships of the Yersinia Plagues

The origins of the three great *Y. pestis* plagues and the reason for such high levels of mortality in the past have long remained unclear. Also uncertain was whether all three pandemic strains arose from the same ancestral lineage or whether each had independently emerged from distinct host reservoirs. While this ambiguity and other questions remain unanswered, such as the genetic properties of *Y. pestis* from animal reservoirs in historical times, phylogenetic analyses of epidemic-associated *Y. pestis* have now begun to shed light on how the three plague epidemics are related to each other.

Strains associated with the first pandemic (the Justinian Plague) collected from two victims buried in a German medieval cemetery, appear on a novel branch that is phylogenetically distinct from strains from the second and third plague pandemics (Figure 3) [37]. The branch has no known contemporary descendants, suggesting that it is either extinct or unsampled in rodent reservoirs [37]. In addition, these Justinian samples were interleaved between two extant groups identified in long-term plague foci in China, which is consistent with the hypothesis that the first pandemic emerged from rodents in Asia and later spread east to Europe [35].

The most infamous of the plague pandemics was the second pandemic in the mid-14th century. Plague persisted in Europe for the next 400 years, and several genetic variants of the second pandemic lineage appear to have continuously emerged and occurred simultaneously in multiple countries during later outbreaks [42]. Comparison of bacterial samples from plague victims in Marseille during the last major European outbreak in the 18th century with

both extant and other historical strains revealed that the Marseille lineage is most closely related to the 14th century plague strains. Cui and colleagues' findings also support this, showing a phylogeny characterized by the emergence of several lineages within a relatively short period of time at the beginning of the second pandemic [43]. To explain this pattern, the authors hypothesized that mutation rate varied across lineages, possibly due to the alternation of endemic and epidemic periods, i.e., epidemics result in replication cycles occurring more frequently and the accumulation of higher number of single nucleotide polymorphisms (SNPs) than enzootic phases of the bacterium [43]. Another possible scenario is that a *Y. pestis* strain responsible for the second pandemic became established in a historical, uncharacterized host reservoir and later evolved locally to sustain repeated outbreaks in Europe over four centuries [42].

Routes of Yersinia Plague Dispersal in Europe

The second pandemic lineage of *Y. pestis* appears to have been introduced into Europe from rodent plague foci in Asia on multiple occasions [44]. Phylogenomic analysis reveals that the most ancestral *Y. pestis* lineages may have likely originated in China or the former Soviet Union, with ten waves of transmission and subsequent radiation in Europe, South America, Africa and Southeast Asia occurring over many centuries [45]. While this is based on a limited number of genome samples, and the exact nature and number of intercontinental transmission events may become clearer with additional sampling efforts, this hypothesis is supported by climate data, which show evidence for repeated climate-driven reintroductions into Europe from Asian reservoirs [44]. In addition, phylogenomic analysis shows that the deepest branching lineages are those isolated near the ancient Silk and Tea-Horse trade

routes [43]. The strong geographical clustering of plague strains on the phylogenomic tree appears to parallel known routes of plague transmission and correspond with historical records, such as the timing of major historical maritime transport of humans and cargo, land trade routes and inter-continental expeditions [11,45], but further additional DNA sampling is needed to elucidate dispersal routes of ancient plague.

Virulence of the Yersinia Plague

It remains an intriguing question why plague has been particularly successful in Europe in the past. Through genome sequencing of ancient *Y. pestis* strains from infected humans in Bronze Age Eurasia (~5,800 years ago), it has been determined that the acquisition of the *Yersinia* murine toxin (*ymt*) was of much importance in the epidemiological history of plague, as this allowed for bacterial survival of *Y. pestis* in the flea vector, giving rise to the large scale pandemics observed in the past millennium [46]. Genome sequencing of the first plague pandemic strains also revealed the presence of non-synonymous SNPs in two virulence genes (*ail* that codes for an outside membrane protein and *yopJ* that codes for an effector protein), which may have contributed to the heightened mortality rate during the first pandemic [37].

Using DNA from dental pulp from Black Death victims exhumed from the East Smithfield burial ground in London, England, Bos and colleagues reported that contemporary strains comprising two main lineages of *Y. pestis* originate from the 14th century epidemic lineage [22]. Further comparison of the genomes of the reconstructed ancient strain and 18 modern strains reveal remarkably few genetic differences in virulence determinants between them over the last 660 years. The authors suggest that the pathogen's severity in medieval times

253 may not have been due to fixed mutations conferring a unique phenotype, but instead may
254 have been caused by non-genetic factors such as climate, host susceptibility, vector dynamics
255 and social conditions. These results were further confirmed by comparing the same ancient *Y*.
256 *pestis* genome with SNP data from a global collection of 289 strains [47].
257
258 While studies of ancient plague samples have given insights into the geographical spread and
259 potential drivers of its success, it is prudent to account for the effect of poor and biased
260 sampling on phylogenetic inference. The lack of samples from a particular geographic area
261 may underestimate or miss entirely its role in further propagating *Y*. *pestis*. In addition, the
262 absence of archeological rodent samples could substantially affect the conclusions we make
263 about the role of animal reservoirs in spreading the disease. Moving forward, it is expected
264 that developments in sampling and recovery of ancient genetic material will allow further
265 studies at the population level and on demographic processes.
266
267
268 **Discovery of Previously Unrecognized Animal Reservoirs**
269 The structure of a phylogenetic tree, particularly in the relationship between historical and
270 extant strains, provides much information regarding the transmission and evolutionary
271 dynamics of the pathogen. Location of particular strains within a tree can provide novel
272 insights on the history of disease transmission and animal reservoirs, as has been shown in
273 the history of *Mycobacterium tuberculosis* in the Americas. Tuberculosis became highly
274 prevalent in Europe and North America in the 18^{th}-19^{th} century and then declined thereafter
275 [1]. In South America, three 1000-year-old samples of *M*. *tuberculosis* from Peruvian

skeletons were sequenced and found to share a more recent common ancestor with pathogens collected from seals and sea lions than with contemporary tuberculosis strains from humans [33]. The most recent common ancestor (MRCA) was estimated to have occurred 4,000-5,000 years ago, providing evidence against introduction of tuberculosis to the Americas via early European contact or human migration across the Bering land bridge, and that the disease may have been introduced via exploitation of marine mammals on the Pacific coast [33]. This study demonstrates how genomic sequencing of ancient pathogens can generate new hypotheses to explain the epidemiological history of a pathogen, in this case the role of previously unrecognized animal reservoirs in disease transmission.

Migration and Transmission of Ancient Pathogens

Genomic diversity is particularly useful in distinguishing distinct bacterial lineages that existed in the past, which can provide insights into human migration and transmission. Such is the case for obligate pathogens such as *M. tuberculosis* [48] and *M. leprae* [49], which can persist in the human host for decades. Leprosy (caused by *Mycobacterium leprae*) was highly prevalent in Europe in the 12^{th}-14^{th} century, and spread to Africa and the Americas in the 16^{th}-17^{th} century [1,49]. It is a chronic disease with potentially debilitating neurological consequences, and the resulting disfigurement of the face and hands has been accompanied by social stigma and exclusion in the past [1]. Comparison of genomic sequences of *M. leprae* has shown that this pathogen has undergone reductive evolution, in which the genome downsized and accumulated a large number of pseudogenes, or inactivated non-functional genes [49]. While this substantial gene decay is expected to give rise to substantial genetic

299 variability, analysis of SNPs and variable number tandem repeats (VNTRs) demonstrated
300 only 807 polymorphic sites across all strains analyzed [31]. Despite very few genetic
301 differences between strains, rare SNPs in the genome divided all extant isolates into five
302 phylogenetic groups, each strongly associated with geographical origins. It was previously
303 thought that type 3 strains, almost exclusively from Europe, disseminated from type 2 strains
304 from Asia and the Middle East with human migration [49]. However, genome sequencing of
305 additional archaeological samples from a medieval leprosy hospital in England showed that
306 type 2 strains were possibly co-dominant or even predominant in Europe during the 11th
307 century [50].
308
309 Not only do ancient pathogens provide insight into the presence of multiple lineages within a
310 species, it also allows us to examine the various genotypes present in a single individual
311 [34,51]. Using metagenomic approaches, Kay *et al*. found that five out of eight bodies
312 recovered from 18th century Hungary yielded more than one *Mycobacterium tuberculosis*
313 genotype [34]. This observation reflects historical and modern differences in tuberculosis
314 epidemiology since multi-strain infections may have potentially been more common in
315 Europe at time of peak tuberculosis compared to contemporary infections [34]. In addition,
316 the same study provided the first evidence of a historical epidemiological link between
317 infections present in two individuals. They found that the same *M*. *tuberculosis* genotypes in
318 samples from both bodies belonged to a mother and her child, supporting either within-
319 family transmission or infection from a common source [34]. Furthermore, varying
320 proportions of the different genotypes found in the samples may suggest spatial heterogeneity
321 in strain distribution [34]. These findings demonstrate that characterization of ancient

pathogen samples can also be useful in providing in-depth information about genetic diversity within a single host and potential transmission.

Genomic Changes Over Time: Comparing Historical and Contemporary Lineages

Comparative genomics of ancient and extant strains allows us to trace specific genetic changes that have occurred over time in pathogenic species and understand the evolutionary relationships of historical and contemporary lineages. Genetic changes through mutations or the horizontal acquisition of loci associated with virulence or transmission may be partially responsible for differences in the severity of diseases today and in the past. This information is also particularly relevant in quantifying the extent to which pathogen genetics may have been associated with devastating levels of mortality in the past, relative to other factors, such as host population and environmental dynamics.

Long-term tracing of genetic adaptation and accurate quantification of rates of evolutionary change are highly informative in understanding how a pathogen becomes more virulent or transmissible, and thus provide insights into how we can effectively manage future epidemics. Genomic comparison of DNA from 11^{th}-14^{th} century skeletons of five individuals who died from leprosy in the United Kingdom, Denmark and Sweden with 11 modern strains of *M. leprae* from multiple leprosy-endemic regions indicates a remarkably high genomic conservation (only 755 SNPs and 57 <7bp-indels) during the last 1,000 years of its evolution [31]. In fact, a total of only 28 nucleotide changes differentiate a closely related pair of modern (from USA) and medieval (from UK) strains. These results are further confirmed by

345 the genome sequencing of two ancient *M. leprae* isolates from 10^{th}-12^{th} century England [50].
346 Schuenemann and colleagues suggested that the severity of the disease in the past and its
347 decline in the succeeding centuries might not have been due to changes in the virulence or
348 transmission capacity of the pathogen, but rather may reflect the greater role of the host (such
349 as immune response and improved infection control behavior) and the presence of other
350 diseases in the past. In the case of leprosy, based on samples dating from the Roman period
351 to the 13^{th} century, it has been suggested that leprosy victims, already having a weakened
352 immune system, are likely to be co-infected with opportunistic *M. tuberculosis* or have a
353 resurgence of a latent tuberculosis infection [52]. Hence, co-infection may have contributed
354 to increased mortality due to tuberculosis and consequently to the decline in leprosy in the
355 population [52].
356
357 Fine-scale identification of the genomic changes occurring over a long period of time has
358 also provided new insights into the evolution of the influenza virus. The 1918-1919 Spanish
359 influenza pandemic, which caused around 50 million deaths worldwide [53], appeared to
360 have rapidly spread in three distinct waves during that period in Europe, Asia and Americas
361 [54]. However, it was unclear whether these were all caused by the same virus strain. A
362 curious characteristic of the Spanish flu was that mortality peaked among healthy young
363 adults (20-40 years old), in contrast to other outbreaks/seasons in which the elderly and
364 young children are usually at greater risk [54]. The genome of the 1918 pandemic influenza
365 virus was first characterized from a lung tissues of a victim from Alaska [55]. Identification
366 of mutational changes within the H1N1 influenza lineage in 4,106 genome sequences from
367 1918 (Spanish flu pandemic) to present indicate that most of the changes are non-adaptive

368 and appear to be degenerative [56]. It was estimated that H1N1 emerged in 1893 and by 1918,
369 had already accumulated ~375 fixed mutations since it first appeared in the human
370 population, while contemporary strains have diverged by as much as 15% from the 1918
371 influenza genotype. The authors contend that the historical decline in H1N1 human mortality
372 after 1918 may be best explained by natural viral attenuation through accumulation of
373 degenerative mutations [56]. However, the role of reassortment between strains from
374 different hosts - such as the triple reassortment of lineages from human, bird and swine that
375 gave rise to the 2009 pandemic lineage [57] - remains a crucial aspect of influenza evolution.
376 Understanding further the history, timing and driving factors of genomic reassortment in
377 influenza evolution will help explain the emergence of new epidemic lineages.
378
379 A total of six documented cholera pandemics occurred in the 19th century, with the first
380 pandemic originating in the Indian subcontinent. The 1961-1975 cholera pandemic first
381 appeared in Indonesia and subsequently spread worldwide. Genome sequencing of a bacterial
382 isolate from a cholera patient of the 1849 Philadelphia, USA outbreak (second pandemic)
383 revealed that it contains all major virulence regions and genomic islands found in the
384 classical strain O395, one of the two subtypes of the pandemic-causing serogroup O1 [58].
385 Phylogenetic analysis and temporal estimation with contemporary strains suggest that the
386 first five cholera pandemics appear to have clonally reemerged and were already circulating
387 in the human population and water sources centuries before the pandemic in the 19th century
388 occurred [58].
389

Other studies that compare genomes collected from historical and extant isolates have been insightful. Genome reconstruction of *Brucella* from a 14th century skeleton in Italy show close phylogenetic relatedness to modern-day isolates in the country, suggesting infection of *B. melitensis* from livestock has continuously occurred in the region over many centuries [59]. The genome of *Helicobacter pylori* isolated from a 5,300-year-old mummy is genetically distinct from most strains common in modern Europe, with extant strains belonging to a recombinant hybrid of two strains related to those that circulate in India and North Africa [60]. Temporal analysis of the hepatitis B virus genome from a 16th century mummy in Korea and contemporary strains in East Asia suggests that the ancestral HBV dates back to ~3,000 years ago and was introduced into the country from Japan and/or China [13].

Along with stochastic evolution, pathogens are constantly adapting to multiple selective pressures imposed by the host and the environment. Over time, genomic signatures due to these changes are erased. Additional sampling during the intervening years, i.e., between epidemics, or from local outbreaks will further clarify the long-term rise and fall in the prevalence of a disease and provide clues to the pathogen's virulence and/or transmission.

Concluding Remarks and Future Directions

While genome sequencing of DNA from ancient disease outbreaks promises to provide new insights into the nature of the co-evolution of a pathogen and its human host (see Outstanding Questions), its utility in the intersection of human history and microbiology remains in its infancy. The importance of historically documented burial grounds, artifacts and other relics

cannot be overestimated, as these can provide a more precise temporal constraint to substantiate findings from genomic studies. Ancient genomic data should be viewed as just one component of available information regarding historical outbreaks, and findings should be viewed alongside archaeological findings and historical records.

The relative contributions of environmental factors in past infections also remain to be precisely determined. Human migration, agriculture, climate change, overcrowding in cities, exploration of new continents, and animal domestication are some of the major factors that may have contributed to the emergence and spread of diseases in ancient history. Recently, it has been reported that tropical diseases carried by humans migrating out of Africa may have contributed to the decline of the Neanderthals [61]. Such enormous upheavals in the society and environment are likely to have important consequences on pathogen biology and a genomic signature may be left behind.

Understanding and interpreting historical events and dynamics usually requires examining several sources of historical as well as biological evidence, which may include microscopy, immune detection, serology and direct detection of microbial components [62,63]. One excellent example is the detection of cell wall lipid biomarkers to identify pathogenic mycobacteria from lesions in ancient human skeletons, which was used to infer the presence of human tuberculosis before agricultural domestication [63]. Most sources of evidence will neither conclusively support nor reject a given hypothesis. Genomic data is no different. Considerable uncertainty remains around historical epidemic dynamics inferred from genomic data, and it remains important to examine hypotheses in context with additional

436 archaeological and historical sources. This is particularly important because the causal agent
437 of many ancient diseases remains uncertain. For example, it has been hypothesized that the
438 English Sweat, which occurred in England in five major outbreaks between 1485-1551), and
439 the Picardy Sweat, which caused regular outbreaks in 18th-19th century France, were caused
440 by hantavirus infection based on similarities in disease symptoms and seasonality [64], but
441 this remains uncertain. As the field of historical genomic epidemiology further develops and
442 new samples can be incorporated into existing phylogenetic trees, new insights into ancient
443 outbreaks will continue to emerge.
444
445 Ancient genomes provide a rich source of information about the long-term development of
446 virulence and niche adaptation of a pathogen. This is particularly relevant today, when a
447 diverse array of deadly microbes, such as SARS [65], Ebola virus [9], Zika virus [66] and
448 *Elizabethkingia* (http://www.cdc.gov/elizabethkingia/outbreaks/index.html), which
449 previously have rarely caused illness in humans but have now emerged as important
450 opportunistic pathogens in multiple countries. With these new threats, it is crucial that we
451 utilize as much information in our toolbox as we can. The cyclic characteristics of some
452 diseases means that what we consider as “new” oubreaks are most often the result of the
453 varying patterns of disease circulation throughout our history; hence, the massive scale and
454 significance of past outbreaks should be a constant reminder to what can happen in the future.
455 Emerging outbreaks can be phylogenetically linked to previous and historical outbreaks
456 where data are available, providing some context to the potential dynamics of the
457 contemporary outbreak. Ancient genomes and the long-term dynamics of epidemics, together
458 with the interactions of climate, environment, vectors and hosts, can be used as input for

459 modeling the potential of today's pathogens to become equally successful in the scale of past
460 pandemics. As we continue to develop strategies (such as vaccine development,
461 antimicrobial usage, diagnostics, surveillance and response, global communication, and
462 availability of data) to manage emerging threats in the 21st century and beyond, we can look
463 back at how ancient pathogens have evolved and adapted throughout the history of
464 humankind, how they have influenced the social structure, values and institutions in the past,
465 and how we can use this information to prepare for future encounters with epidemic diseases.
466
467
468 **Text Box**
469
470 **Box 1. Recovery and Analysis of Ancient Genomes: Brief Overview**
471 *Sample Collection*
472 Source material is usually obtained from bone, dental pulp, coprolites (faeces) and
473 mummified tissues of human remains. While some diseases leave skeletal deformities, such
474 as tuberculosis, which can be identified via tuberculous lesions in skeletal remains, most do
475 not. In the case of syphilis, such evidence can be ambiguous, since other treponemal diseases
476 such as yaws and bejel may also present similar skeletal lesions [67], contributing to the
477 difficulty in resolving the origins of venereal syphilis in Renaissance Europe [68,69].
478 Another important consideration is temporal calibration of burial grounds and other
479 archaeological sites, including mass graves, in order to date the samples collected. This could
480 be done through retrieval from written record (e.g., death registers) or radiocarbon dating of
481 ancient corpses, inscriptions and archaeological specimens (e.g., textiles, coin, coffin,

jewelry) [70]. For an excellent discussion of the various techniques used in sampling, verification, microscopy, serological methods, immunodetection, molecular and genomic analyses of ancient pathogens, see the review by [70].

DNA/RNA Recovery and Sequencing

After decontamination procedures, extracted DNA is converted to DNA libraries for genome sequencing. In most cases, no fragmentation step or size selection after end repair is carried out due to the degraded nature and the miniscule proportions of ancient DNA. Some laboratories use an unbiased metagenomic shotgun sequencing method approach (e.g., [59]) to seek out pathogen DNA rather than targeted amplification or capture of the DNA of interest (e.g., [71]). Genomic sequences are then pre-processed (e.g., adapter clipping, quality trimming) and assembled by mapping them to a reference genome, usually a modern-day strain. Multiple sequences are then aligned and variable positions identified.

Phylogenetic and Temporal Analysis

Phylogenetic relationships can be inferred using various tree-building methods. Maximum likelihood methods, such as RAxML [72], assume independence between all sites in the alignment and are commonly used to infer phylogenies from samples of bacterial genomes. A topology is sampled and the likelihood of the data given the fit to the proposed topology and nucleotide substitution model is calculated. The process is iterated and the topology with the highest likelihood is selected. Establishing the timing of internal phylogenetic tree nodes provides evidence for when distinct clades diverged, or when important mutation or gene acquisition events occurred. A common approach is to use BEAST, a software tool which

implements a Bayesian Markov chain Monte Carlo approach to provide posterior estimates of key evolutionary and tree structure parameters [73]. Given an evolutionary and demographic model, the algorithm samples across tree space, providing the distribution for the timing of each internal tree node.

Figure Legends

Figure 1. Overview and Timeline of Historically Notable Disease Outbreaks in Human History

Colored dots represent different outbreaks and epidemics. Black dots indicate disease outbreaks of unresolved origins [1,74–82]. The origin of syphilis remains contentious, with two hypotheses put forward to explain the epidemic in Europe in the 12-14th century [1,68,69,83]. Not shown are seven Bronze Age (~3,000 BC) *Yersinia* plague strains [46]. Location of dots represents approximate time and should not be taken as precise estimates of the time of occurrence of the disease.

Figure 2. Overview of Microbial Genomics of Ancient Pathogens

Studying ancient pathogens can be a valuable tool in answering many historical and biological questions. These include what could have caused the high mortality of past diseases, how they spread across continents, how did vectors and host reservoirs contribute to their virulence and transmission, and the origins and diversity of pathogens in the past. For a

more detailed explanation of the methods used in ancient pathogen genomics, we refer the reader to an excellent review by Drancourt and Raoult [70].

Figure 3. Phylogenetic Relationship of *Y. pestis*

The three *Yersinia* plague pandemics (colored branches) are shown with contemporary *Yersinia* lineages (black branches). The map shows the approximate locations from where the pandemic samples were collected. *Yersinia pseudotuberculosis* was used as outgroup. Locations are abbreviated as CHN (China), FSU (Former Soviet Union), GEO (Georgia), AFR (Africa), MNG (Mongolia), NPL (Nepal), and IRN (Iran). Adapted with permission from [42].

Table Legend

Table 1. Genomes of pathogens collected from archaeological remains sequenced to date

References

1 Sherman, I.T. (2006) *The Power of Plagues*, ASM Press.

2 Watts, S. (1997) *Epidemics and History: Diesease, Power and Imperialism*, Yale University Press.

3 Karamanou, M. *et al*. (2012) From miasmas to germs: a historical approach to theories of infectious disease transmission. *Infez. Med*. 20, 58–62

4 Mendelsohn, J.A. (2002) "Like all that lives": biology, medicine and bacteria in the age of Pasteur and Koch. *Hist. Philos. Life Sci*. 24, 3–36

5 Eppinger, M. *et al*. (2014) Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio* 5, e01721

6 Grad, Y.H. and Waldor, M.K. (2013) Deciphering the origins and tracking the evolution of cholera epidemics with whole-genome-based molecular epidemiology. *mBio* 4, e00670-613

7 Hasan, N.A. *et al*. (2012) Genomic diversity of 2010 Haitian cholera outbreak strains. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 109, E2010-2017

8 Carroll, M.W. *et al*. (2015) Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. *Nature* 524, 97–101

9 Gire, S.K. *et al*. (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 345, 1369–1372

10 Simon-Loriere, E. *et al*. (2015) Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature* 524, 102–104

11 Bossak, B.H. and Welford, M.R. (2009) Did medieval trade activity and a viral etiology control the spatial extent and seasonal distribution of Black Death mortality? *Med*. *Hypotheses* 72, 749–752

12 Taubenberger, J.K. *et al*. (1997) Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science* 275, 1793–1796

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

14 Gilbert, M.T.P. *et al*. (2007) The emergence of HIV/AIDS in the Americas and beyond. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 104, 18566–18570

15 Fordyce, S.L. *et al*. (2013) Deep sequencing of RNA from ancient maize kernels. *PloS One* 8, e50961

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

17 Pääbo, S. *et al*. (2004) Genetic analyses from ancient DNA. *Annu*. *Rev*. *Genet*. 38, 645–679

18 Briggs, A.W. *et al*. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 104, 14616–14621

19 Rohland, N. *et al*. (2015) Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos*. *Trans*. *R*. *Soc*. *Lond*. *B*. *Biol*. *Sci*. 370, 20130624

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

21 Gansauge, M.-T. and Meyer, M. (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat*. *Protoc*. 8, 737–748

22 Bos, K.I. *et al*. (2011) A draft genome of Yersinia pestis from victims of the Black Death. *Nature* 478, 506–510

23 Schubert, M. *et al*. (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat*. *Protoc*. 9, 1056–1082

24 Peltzer, A. *et al*. (2016) EAGER: efficient ancient genome reconstruction. *Genome Biol*. 17, 60

25 Wood, J.W. *et al*. (1992) The Osteological Paradox: Problems of Inferring Prehistoric Health from Skeletal Samples [and Comments and Reply]. *Curr*. *Anthropol*. 33, 343–370

26 Bouwman, A.S. *et al*. (2012) Genotype of a historic strain of Mycobacterium tuberculosis. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 109, 18511–18516

27 Adler, C.J. *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, 455e1

28 Warinner, C. *et al*. (2014) Pathogens and host immunity in the ancient human oral cavity. *Nat*. *Genet*. 46, 336–344

29 Campana, M.G. *et al*. (2014) False positives complicate ancient pathogen identifications using high-throughput shotgun sequencing. *BMC Res*. *Notes* 7, 111

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

31 Schuenemann, V.J. *et al*. (2013) Genome-wide comparison of medieval and modern Mycobacterium leprae. *Science* 341, 179–183

32 Rajaraman, A. *et al*. (2013) FPSAC: fast phylogenetic scaffolding of ancient contigs. *Bioinforma*. *Oxf*. *Engl*. 29, 2987–2994

33 Bos, K.I. *et al*. (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514, 494–497

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

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

36 Little, L.K. (2007) *Plague and the End of Antiquity: the Pandemic of 541-750*, Cambridge Univ. Press.

37 Wagner, D.M. *et al*. (2014) Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. *Lancet Infect*. *Dis*. 14, 319–326

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

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

40 Ben-Ari, T. *et al*. (2012) Identification of Chinese plague foci from long-term epidemiological data. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 109, 8196–8201

41 Zhang, Z. *et al*. (2009) Spatial variation of Yersinia pestis from Yunnan Province of China. *Am*. *J*. *Trop*. *Med*. *Hyg*. 81, 714–717

42 Bos, K.I. *et al*. (2016) Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife* 5,

43 Cui, Y. *et al*. (2013) Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 110, 577–582

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

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

640 46 Rasmussen, S. *et al*. (2015) Early divergent strains of Yersinia pestis in Eurasia 5,000
641 years ago. *Cell* 163, 571–582
642 47 Bos, K.I. *et al*. (2012) Yersinia pestis: new evidence for an old infection. *PloS One* 7,
643 e49803
644 48 Mokrousov, I. *et al*. (2016) Latin-American-Mediterranean lineage of Mycobacterium
645 tuberculosis: Human traces across pathogen's phylogeography. *Mol*. *Phylogenet*. *Evol*.
646 99, 133–143
647 49 Monot, M. *et al*. (2005) On the origin of leprosy. *Science* 308, 1040–1042
648 50 Mendum, T.A. *et al*. (2014) Mycobacterium leprae genomes from a British medieval
649 leprosy hospital: towards understanding an ancient epidemic. *BMC Genomics* 15, 270
650 51 Müller, R. *et al*. (2014) Genotyping of ancient Mycobacterium tuberculosis strains
651 reveals historic genetic diversity. *Proc*. *Biol*. *Sci*. 281, 20133236
652 52 Donoghue, H.D. *et al*. (2005) Co-infection of Mycobacterium tuberculosis and
653 Mycobacterium leprae in human archaeological samples: a possible explanation for the
654 historical decline of leprosy. *Proc*. *Biol*. *Sci*. 272, 389–394
655 53 Johnson, N.P.A.S. and Mueller, J. (2002) Updating the accounts: global mortality of the
656 1918-1920 "Spanish" influenza pandemic. *Bull*. *Hist*. *Med*. 76, 105–115
657 54 Taubenberger, J.K. and Morens, D.M. (2006) 1918 Influenza: the mother of all
658 pandemics. *Emerg*. *Infect*. *Dis*. 12, 15–22
659 55 Tumpey, T.M. *et al*. (2005) Characterization of the reconstructed 1918 Spanish influenza
660 pandemic virus. *Science* 310, 77–80

56 Carter, R.W. and Sanford, J.C. (2012) A new look at an old virus: patterns of mutation accumulation in the human H1N1 influenza virus since 1918. *Theor. Biol. Med. Model.* 9, 42

57 Lam, T.T.Y. *et al*. (2011) Reassortment events among swine influenza A viruses in China: implications for the origin of the 2009 influenza pandemic. *J. Virol.* 85, 10279–10285

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

59 Kay, G.L. *et al*. (2014) Recovery of a medieval Brucella melitensis genome using shotgun metagenomics. *mBio* 5, e01337-1314

60 Maixner, F. *et al*. (2016) The 5300-year-old Helicobacter pylori genome of the Iceman. *Science* 351, 162–165

61 Houldcroft, C.J. and Underdown, S.J. (2016) Neanderthal genomics suggests a pleistocene time frame for the first epidemiologic transition. *Am. J. Phys. Anthropol.*

62 Tran, T.-N.-N. *et al*. (2011) Beyond ancient microbial DNA: nonnucleotidic biomolecules for paleomicrobiology. *BioTechniques* 50, 370–380

63 Baker, O. *et al*. (2015) Human tuberculosis predates domestication in ancient Syria. *Tuberc. Edinb. Scotl.* 95 Suppl 1, S4–S12

64 Heyman, P. *et al*. (2014) Were the English sweating sickness and the Picardy sweat caused by hantaviruses? *Viruses* 6, 151–171

65 Cherry, J.D. (2004) The chronology of the 2002-2003 SARS mini pandemic. *Paediatr. Respir. Rev.* 5, 262–269

683 66 Faria, N.R. *et al*. (2016) Zika virus in the Americas: Early epidemiological and genetic
684 findings. *Science* 352, 345–349
685 67 Rothschild, B.M. and Rothschild, C. (1995) Treponemal disease revisited: skeletal
686 discriminators for yaws, bejel, and venereal syphilis. *Clin*. *Infect*. *Dis*. *Off*. *Publ*. *Infect*.
687 *Dis*. *Soc*. *Am*. 20, 1402–1408
688 68 Erdal, Y.S. (2006) A pre-Columbian case of congenital syphilis from Anatolia (Nicaea,
689 13th century AD). *Int*. *J*. *Osteoarchaeol*. 16, 16–33
690 69 Harper, K.N. *et al*. (2011) The origin and antiquity of syphilis revisited: an appraisal of
691 Old World pre-Columbian evidence for treponemal infection. *Am*. *J*. *Phys*. *Anthropol*.
692 146 Suppl 53, 99–133
693 70 Drancourt, M. and Raoult, D. (2005) Palaeomicrobiology: current issues and perspectives.
694 *Nat*. *Rev*. *Microbiol*. 3, 23–35
695 71 Schuenemann, V.J. *et al*. (2011) Targeted enrichment of ancient pathogens yielding the
696 pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proc*. *Natl*. *Acad*. *Sci*.
697 *U*. *S*. *A*. 108, E746-752
698 72 Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-
699 analysis of large phylogenies. *Bioinforma*. *Oxf*. *Engl*. 30, 1312–1313
700 73 Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by
701 sampling trees. *BMC Evol*. *Biol*. 7, 214
702 74 Adamson, P.B. (1976) Schistosomiasis in Antiquity. *Med*. *Hist*. 20, 176–188
703 75 Cunha, B.A. (2004) The cause of the plague of Athens: plague, typhoid, typhus, smallpox,
704 or measles? *Infect*. *Dis*. *Clin*. *North Am*. 18, 29–43

76 Cunha, C.B. (2007) Prolonged and perplexing fevers in antiquity: malaria and typhoid fever. *Infect. Dis. Clin. North Am.* 21, 857–866, vii

77 Gilliam, J.F. (1961) The Plague under Marcus Aurelius. *Am. J. Philol.* 82, 225–251

78 Littman, R.J. and Littman, M.L. (1973) Galen and the Antonine Plague. *Am. J. Philol.* 94, 243–255

79 Littman, R.J. (2009) The plague of Athens: epidemiology and paleopathology. *Mt. Sinai J. Med. N. Y.* 76, 456–467

80 Papagrigorakis, M.J. *et al.* (2006) DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens. *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.* 10, 206–214

81 Retief, F.P. and Cilliers, L. (1998) The epidemic of Athens, 430-426 BC. *South Afr. Med. J. Suid-Afr. Tydskr. Vir Geneeskd.* 88, 50–53

82 Trevisanato, S.I. (2007) The "Hittite plague", an epidemic of tularemia and the first record of biological warfare. *Med. Hypotheses* 69, 1371–1374

83 Harper, K.N. *et al.* (2008) On the origin of the treponematoses: a phylogenetic approach. *PLoS Negl. Trop. Dis.* 2, e148

Outstanding Questions

- How might we utilize technological advances in sequencing technology to improve the quality and feasibility of sequencing ancient pathogens?
- What genomic and ecological factors could have driven the emergence of the seven cholera pandemics, and how is each pandemic lineage related to one another?
- Are there other animal reservoirs that may have existed in the past and contributed to the geographic spread of ancient plague and other zoonotic diseases?
- How did the population structure and dynamics of ancient leprosy and tuberculosis change over the course of the pandemic (initial outbreak, decline and/or resurgence) and across different geographical areas?

Figure 1

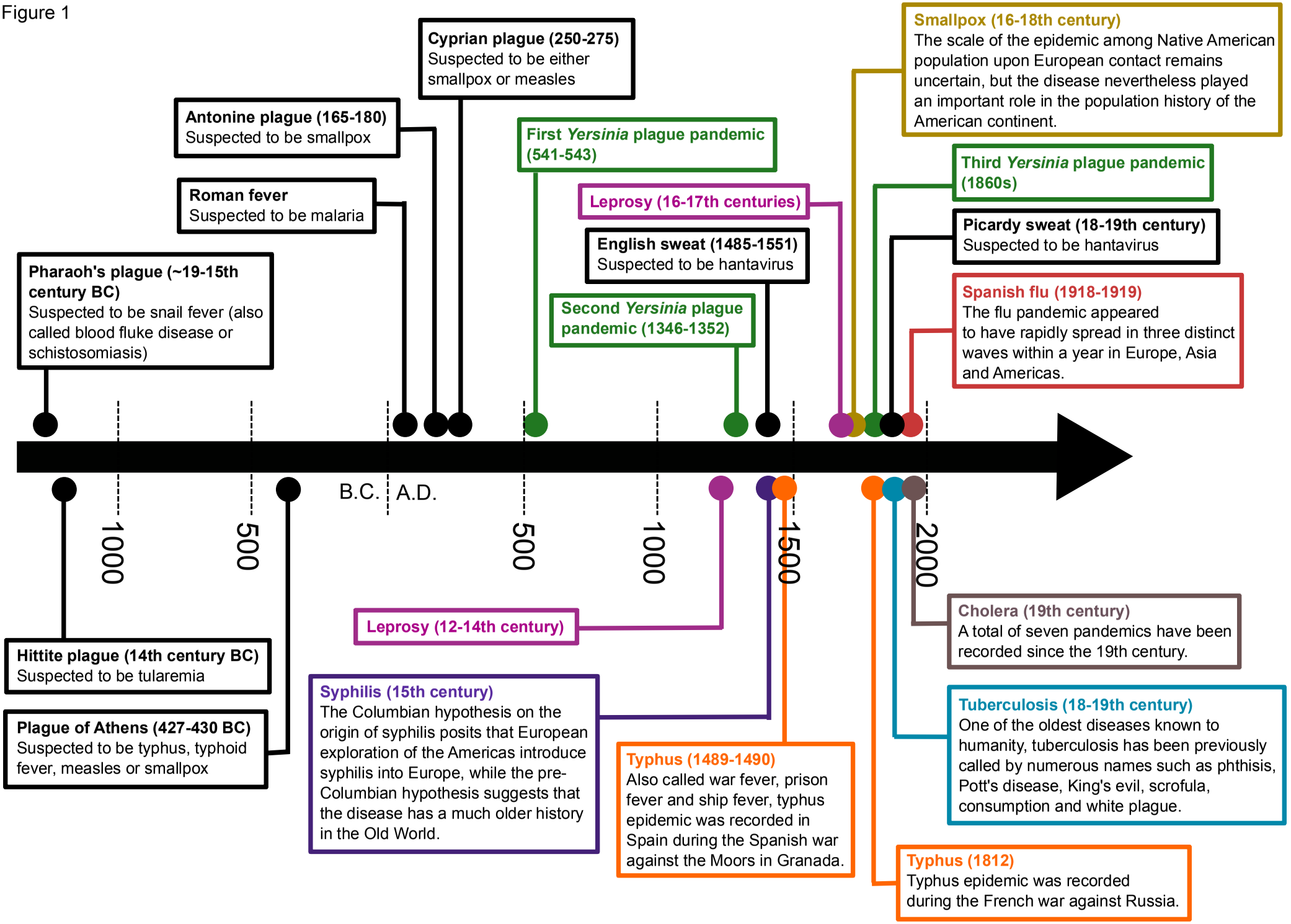
Smallpox (16-18th century)
The scale of the epidemic among Native American population upon European contact remains uncertain, but the disease nevertheless played an important role in the population history of the American continent.
Cyprian plague (250-275)
Suspected to be either smallpox or measles
Antonine plague (165-180)
Suspected to be smallpox
First Yersinia plague pandemic (541-543)
Third Yersinia plague pandemic (1860s)
Roman fever
Suspected to be malaria
Leprosy (16-17th centuries)
Picardy sweat (18-19th century)
Suspected to be hantavirus
English sweat (1485-1551)
Suspected to be hantavirus
Pharaoh's plague (~19-15th century BC)
Suspected to be snail fever (also called blood fluke disease or schistosomiasis)
Second Yersinia plague pandemic (1346-1352)
Spanish flu (1918-1919)
The flu pandemic appeared to have rapidly spread in three distinct waves within a year in Europe, Asia and Americas.
B.C.
A.D.
1000
500
500
1000
1500
2000
Leprosy (12-14th century)
Cholera (19th century)
A total of seven pandemics have been recorded since the 19th century.
Hittite plague (14th century BC)
Suspected to be tularemia
Syphilis (15th century)
The Columbian hypothesis on the origin of syphilis posits that European exploration of the Americas introduce syphilis into Europe, while the pre-Columbian hypothesis suggests that the disease has a much older history in the Old World.
Tuberculosis (18-19th century)
One of the oldest diseases known to humanity, tuberculosis has been previously called by numerous names such as phthisis, Pott's disease, King's evil, scrofula, consumption and white plague.
Plague of Athens (427-430 BC)
Suspected to be typhus, typhoid fever, measles or smallpox
Typhus (1489-1490)
Also called war fever, prison fever and ship fever, typhus epidemic was recorded in Spain during the Spanish war against the Moors in Granada.
Typhus (1812)
Typhus epidemic was recorded during the French war against Russia.

Figure 2

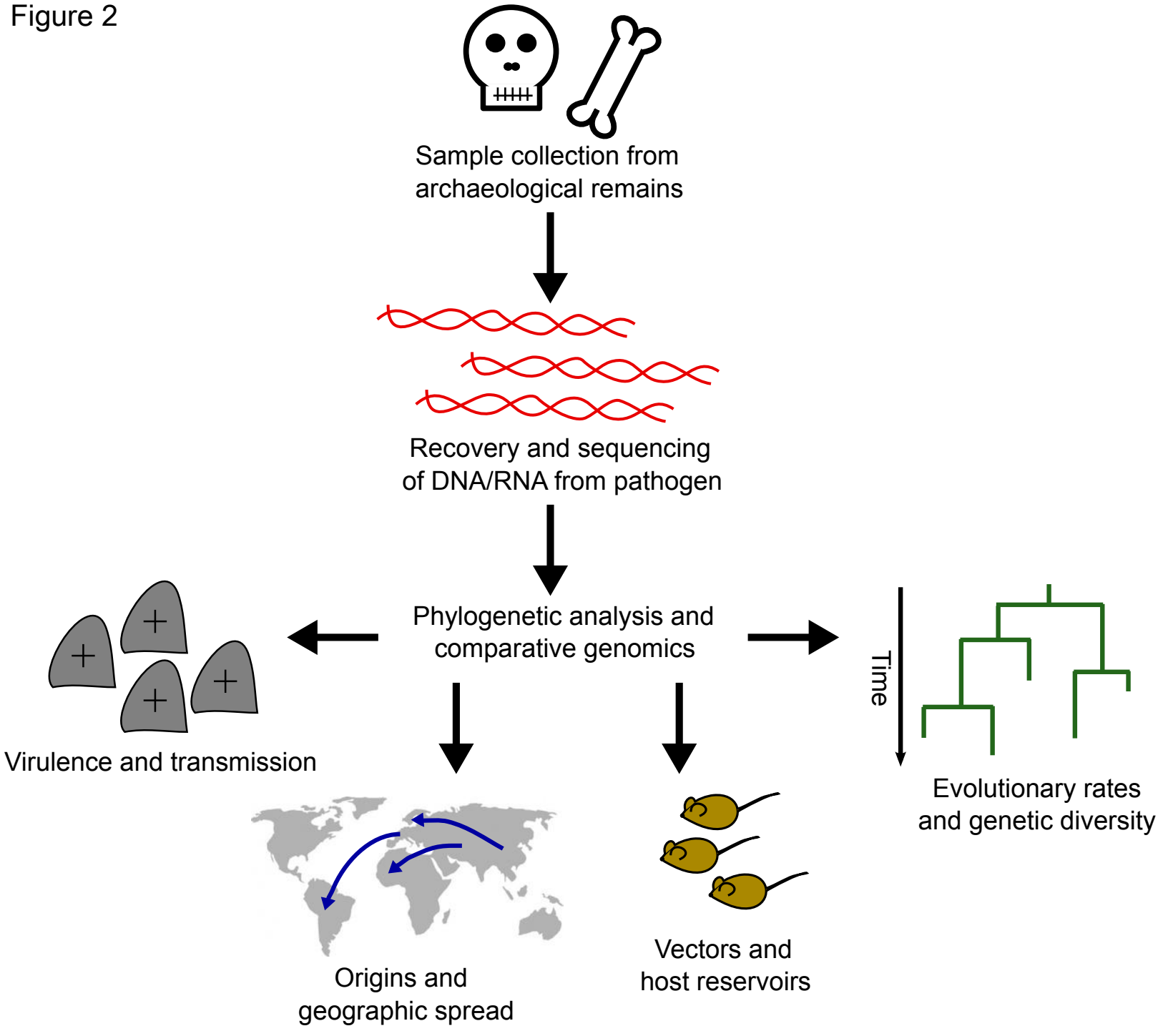
Sample collection from archaeological remains
Recovery and sequencing of DNA/RNA from pathogen
Phylogenetic analysis and comparative genomics
Virulence and transmission
Origins and geographic spread
Vectors and host reservoirs
Time
Evolutionary rates and genetic diversity

Figure 3

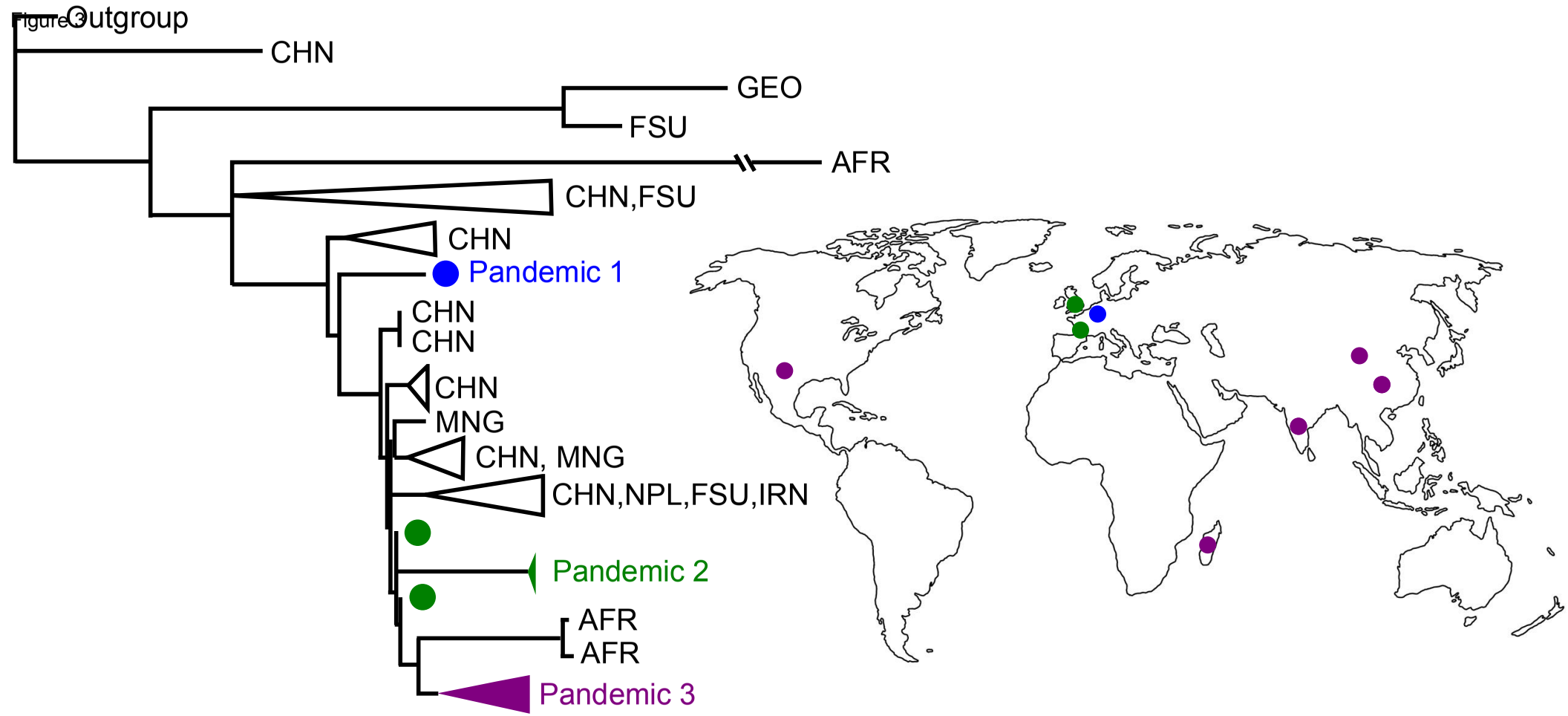
Outgroup
CHN
GEO
FSU
AFR
CHN,FSU
CHN
Pandemic 1
CHN
CHN
CHN
MNG
CHN, MNG
CHN,NPL,FSU,IRN
Pandemic 2
AFR
AFR
Pandemic 3

Table 1. Genomes of pathogens collected from archaeological remains sequenced to date

Pathogen	Number of sequenced genomes	Reference
Brucella melitensis	1	[59]
Helicobacter pylori	1	[60]
Mycobacterium leprae	7	[27, 28]
Mycobacterium tuberculosis	18	[25, 29, 49]
Vibrio cholerae	1 (Second pandemic)	[37]
Yersinia pestis	7 (Bronze Age)	[22]
	2 (First pandemic)	[12]
	7 (Second pandemic)	[17, 23]
Influenza virus (Spanish flu)	1	[34]
Hepatitis B virus	4	[13]