Ancient pathogen genomics: insights into timing and adaptation

Kelly M. Harkins 1,*, Anne C. Stone

Center for Bioarchaeological Research, School of Human Evolution and Social Change, Arizona State University, United States

A R T I C L E   I N F O

Article history:
Received 13 February 2014
Accepted 5 November 2014
Available online 18 December 2014

Keywords:
Ancient DNA
Pathogen evolution
Human disease

A B S T R A C T

Disease is a major cause of natural selection affecting human evolution, whether through a sudden pandemic or persistent morbidity and mortality. Recent contributions in the field of ancient pathogen genomics have advanced our understanding of the antiquity and nature of human-pathogen interactions through time. Technical advancements have facilitated the recovery, enrichment, and high-throughput sequencing of pathogen and parasite DNA from archived and archaeological remains. These time-stamped genomes are crucial for calibrating molecular clocks to infer the timing of evolutionary events, while providing finer-grain resolution to phylogenetic reconstructions and complex biogeographical patterns. Additionally, genome scale data allow better identification of substitutions linked to adaptations of the pathogen to their human hosts. As methodology continues to improve, ancient genomes of humans and their diverse microbiomes from a range of eras and archaeological contexts will enable population-level ancient analyses in the near future and a better understanding of their co-evolutionary history.

© 2014 Elsevier Ltd. All rights reserved.

Introduction

A driving force behind the field of ancient pathogen genomics is the necessity to characterize how human interactions with commensal organisms, pathogens, and hosts impact human evolution. Today, populations and individuals differ in susceptibility to disease as a direct result of our evolutionary history with pathogens (Gomez et al., 2014; Karlsson et al., 2014). Likewise, pathogens evolve in response to human biological change alongside sociocultural and technological developments. For decades, archaeological, historic, and modern molecular approaches have offered independent evidence to address questions about health, disease, and adaptation to pathogens in the recent and deep past. Rapidly advancing next-generation sequencing (NGS) technologies adapted for archaeological or archived samples (e.g., Hodges et al., 2009; Burbano et al., 2010; Maricic et al., 2010; Meyer and Kircher, 2010; Kircher et al., 2011a) have revolutionized the field, reshaping our understanding of pathogen origins and evolution, and the historical and cultural processes that are central to contextualizing disease in past human groups.

The inaugural studies identifying ancient pathogen DNA from archaeological remains (Spigelman and Lemma, 1993; Salo et al., 1994) suffered from the limitations common to pioneering efforts (Pääbo et al., 2004; Willerslev and Cooper, 2006), and the field remains in an intense period of discovery. Criticism directed at the inevitable oversights of early paleogenetic research, which mostly employed traditional polymerase chain reaction (PCR), has resulted in sweeping improvements in the field by way of stringent laboratory practices and authentication methods that range from technical to computational (Cooper and Poinar, 2000; Willerslev and Cooper, 2005; Stone et al., 2009; Jónsson et al., 2013; Gansauge and Meyer, 2014; Skoglund et al., 2014).

The challenges of sequencing ancient DNA (aDNA) stem from observations that DNA molecules preserved within archaeological or historic remains (e.g., bone, teeth, paraffin-embedded or other tissue, coprolites, botanical, or archived material) are fragmented, damaged, and, with rare exception, dominated by 99% or more of contaminating microbial DNA sequences. The application of NCS to aDNA massively parallelizes the sequencing of short, degraded molecules in unprecedented read depths, enabling characterization of degradation patterns expected of ancient molecules (Briggs et al., 2007) and more accurate alignments and confidence in nucleotide determination at a given site. Nevertheless, the notorious difficulties of authenticating ancient sequence data and discriminating false-positives are by no means eliminated by the NGS revolution (Campana et al., 2014), especially when the target of interest is a pathogen and extant microbial diversity is largely unknown (Rappe and Giovannoni, 2003). The complex microbial profiles common to
archaeological bone and other ancient sources are only beginning to be characterized (Der Sarkissian et al., 2014). Next-generation sequencing techniques have, however, mitigated some well-known problems of using PCR to detect ancient pathogen DNA (Drancourt and Raoult, 2005; Roberts and Ingham, 2008), particularly by generating DNA damage profiles, while allowing the field to move beyond simply confirming that a pathogen is present.

While confronting these obstacles, ancient pathogen research contributes to our understanding of infectious disease evolution by providing time-stamped sequence data to integrate into phyloge-netic reconstructions and to use as crucial calibration points to estimate the timing of divergence events, thus testing long held hypotheses regarding the extent of our coevolutionary history with pathogenic and commensal organisms. In addition to providing finer-grain resolution to phylogenetic inferences and complex biogeographical patterns, genome scale data are enabling the identification of substitutions linked to coevolutionary adaptations of the pathogen to their human hosts. Over a decade of technical improvements in genomics has provided a foundation to begin investigating the less-chartered realm of microbes. Thus, the nascent stages of ancient pathogen genomics have begun, often with surprising results. Most notably this early research has focused on many human infections, including the 1918 ‘Spanish’ influenza virus (Taubenberger et al., 2005), HIV-1 virus (Worobey et al., 2008), plague (Ros et al., 2011; Schuenemann et al., 2011; Achtan, 2012), leprosy (Schuenemann et al., 2013), tuberculosis (Bouwman et al., 2012; Chan et al., 2013; Ros et al., 2014), and, most recently, cholera (Devault et al., 2014). Additionally, the major impact that crop diseases have had on human groups throughout history highlights plant pathogens as another interesting avenue of paleogenetic research (e.g., Ristaino et al., 2001; Smith et al., 2014). Overall, this review examines recent contributions of ancient pathogen genomics to our understanding of disease origins, as well as the evolutionary and biocultural processes affecting human-pathogen interaction and adaptation.

**A starting point: the ‘Paleolithic baseline’**

One of the most influential developments in human history—the intensification of agriculture and domestication—also carried with it enduring consequences for human health. This shift, facilitated by an increase in population size, density, fertility, and eventually, expansion, is referred to as the Neolithic Demographic Transition (NDT). The NDT is thought to have markedly contrasted with the longstanding population structure and demographic patterns of hominin foragers, thus representing one of the most fundamental structural processes in our history (Bocquet-Appel, 2002, 2011; Bocquet-Appel and Bar-Yosef, 2008).

The current understanding of hunter-gatherer population structure forms a ‘baseline’ (or what might be called the ‘Paleolithic baseline’) from which many evaluate health and disease, and develop hypotheses about the nature of human-pathogen interaction in pre- and post-Neolithic society (Armelagos et al., 2005). The immense shift in subsistence and demographic occurring at the NDT has caused unintended side effects for human health and consequently has been regarded as the ‘first epidemiological transition’ (Armelagos et al., 1996; Dobson and Carper, 1996; Moodley et al., 2012). Epidemological transitions focus on trends in disease prevalence and mortality, and there are inherent theoretical and methodological problems generalizing these trends from observations in modern hunter-gatherer groups (Lewin, 1988; Barrett et al., 1988) or from archaeological skeletal collections, especially in communities expected to be expanding during the NDT (Wood et al., 1992). However, as scholars of the twentieth and twenty-first centuries contemplate the origins of infectious disease, many have focused upon this demographic and epidemiological transition to examine the processes and conditions that contribute to the persistence of old or emergence of new pathogens within specific spatiotemporal contexts.

Modern and ancient pathogen genome data are, however, suitable for addressing some key assumptions inherent in the ‘Paleolithic baseline’ and thus those associated with the first epidemiological transition. These longstanding assumptions established expectations for the types and attributes of pathogens supportable within small, hunter-gatherer communities in contrast to the emergence of highly virulent, density-dependent, zoonotic diseases in post-agricultural groups (see Burnet, 1962; Cockburn, 1967), resulting in the classification of ‘new’ or ‘old’ pathogens that can now be examined explicitly with genetic data. Of course it may be overly simplistic to assume an isomorphic relationship between population structure/subsistence strategy and patterns of disease. Environment (e.g., climate, rainfall), ecology (e.g., number of bird and mammal species), and socioeconomic factors in a region contribute considerably to human pathogen richness. However, as a general trend, the range of diseases as well as the sustained transmission of viruses and bacteria is limited in smaller, less dense foraging communities (Nunn et al., 2003; Jones et al., 2008; Dunn et al., 2010). Although the transition to and from the Mesolithic period and its nuances largely are overlooked, this ‘Paleolithic baseline’ has proven useful for addressing assumptions concerned with the antiquity of pathogens.

Current ancient pathogen research continues to ask: Which diseases are, in fact, old? Which diseases did humans encounter in the New World, and which did they carry with them? Do pandemics signify human exposure to a ‘new’ pathogen or genetic adaptation on the part of the pathogen? Is the long-held ‘conventional wisdom’ that virulence is a consequence of a long co-evolutionary history in a host species true? Both timing and genome scale data are crucial to examining these outstanding questions.

**Fitting human-pathogen coevolution to a temporal framework**

Timing is critical to understanding the emergence of a human disease and is essential for testing hypotheses about human interaction with pathogens in the course of our evolutionary history. Ancient DNA sequences with time estimates from archaeological/historical contexts or radiocarbon dates can provide calibrations for estimating substitution rates and divergence times at important phylogenetic nodes (Rambaut, 2000; Drummond et al., 2002). Ancient and archived samples can also resolve inconsistencies when timing of older divergence events is underestimated (Wertheim and Pond, 2011). Although calibrating evolutionary clocks is notoriously difficult, current methods account for temporally sampled sequences and attempt to model the uncertainty in sample ages, which pose problems for phylogenetic and dating reconstructions (Ho et al., 2005, 2007, 2011; Ho and Larson, 2006; Debyrawe and Poinar, 2009; Ho and Phillips, 2009). Sample-dating errors, however, appear to infect only minor effects on substitution rate and dating estimates, a conclusion that Molak et al. (2013) extrapolate further to include errors inherent in calibrating radiocarbon dates to calendar ages. A restricted number of ancient sequences in their dataset, however, did impact root age estimates negatively. This observation is the most problematic for current ancient genetics research, considering the difficulties of obtaining multiple ancient sequences with sufficient quality and variation to incorporate into phylogenetic analyses. Population- or community-level ancient analyses are not currently the standard, but given the current trajectory of the field, the number of ancient sequences may no longer be a limiting factor in the near future.
RNA or single-stranded DNA (ssDNA) viruses are ideal for studying rates of evolutionary change because of their fast mutation rate (Duffy et al., 2008) and because they change on a scale such that they are ‘measurably evolving’ (Drummond et al., 2003). These features allow researchers to use samples collected over relatively short time periods, for example tissue samples from infected patients, to calibrate molecular clocks and test models of evolution (Homes, 2008; Shapiro et al., 2011). Research on HIV evolution demonstrates the importance of historically-sampled sequences for rate calibration to help define the timescale of evolutionary change, and, in turn, the timing of divergence events in phylogenies. Methodology specifically designed for recovering degraded and damaged nucleic acid (e.g., Rohland and Hofreiter, 2007; Dabney et al., 2013a) and phylogenetics programs designed to integrate heterochronous data (Anderson et al., 2005; Drummond and Rambaut, 2007; Yang et al., 2007) have benefitted research in viral evolution. Before the AIDS epidemic, many hospitals in west-central Africa collected Bouin’s-fixed paraffin-embedded samples from patients. Ancient DNA extraction protocols facilitated the isolation of genetic information from a sample, DRC60, archived in 1960. The resulting sequence information coupled with a previously successful isolation from a human blood sample, ZR.59, collected in 1959 (Korber et al., 2000) allowed Worobey et al. (2008) to make a number of significant conclusions, namely that HIV-1 evolves in a clock-like fashion and that high sequence diversity within the region, in contrast to founder effects seen outside, is consistent with the hypothesis that central Africa is the site of origin.

In addition to using dates from time-stamped samples, dates inferred from biogeographical patterns represent another tool to calibrate phylogenetic analyses and convert evolutionary distance into time. Worobey et al. (2010) employ this strategy to address similar questions of rate evolution and issues of dating deeper time scales in simian immunodeficiency virus (SIV). Simian immunodeficiency virus is widespread throughout African non-human primates, suggesting an ancient origin for the virus. This is in contrast to molecular clocks calibrated with the HIV sequences sampled from the 1950s and 1960s that indicate an origin of SIV that is just centuries old (Wertheim and Worobey, 2009). To address this discrepancy, the authors sampled SIV in non-human primates on Bioko Island, a landmass known to have separated from mainland Africa ca. 10–12 kya (thousands of years ago). Worobey et al.’s (2010) identify the misleading effect of using recent tip dates to calibrate deeper nodes by showing that the previous rate estimates are ~125 x faster than those calibrated with Bioko landmass separation. Furthermore, their results demonstrate that the SIV phylogeny is on the order of thousands, not millions (or hundreds) of years old and adjust our estimates of when SIV ‘jumped’ into humans and evolved into HIV.

Generally, sampling of non-human primates, as with human remains, is often difficult but worthwhile, given that non-human primates are known to harbor the closest-known relatives of many human infectious diseases (Wolfe et al., 1998). To date, there is no systematic effort to monitor pathogens emerging from animals (Wolfe et al., 2007). For this reason, very little is known about wild or natural reservoirs, or the effect of viruses and other pathogens on fitness costs across non-human primate taxa. Although Wertheim and Worobey (2009) state that SIV is generally benign in its natural host, recent research demonstrates fitness costs in chimpanzees (Keele et al., 2009). Such observations are important because symptoms of disease may not mimic those of humans, but in the case of HIV/AIDS they may have a health burden with evolutionary implications. Much remains unknown regarding the fitness costs of pathogens in natural hosts or reservoirs, and considerable scholarship and debate is devoted to understanding the relationship of pathogen virulence to long term host specificity or coevolution (see Lenski and May, 1994; Read, 1994; Alizon et al., 2009). This research underscores the importance of systematically sampling natural hosts, including domestic animals (both modern and ancient where available), for future research on the evolution of infectious disease.

Timing is not only important for understanding zoonotic events in the past, but it also informs us about significant human transmission events, such as the emergence of pandemics. For example, Hepatitis C virus (HCV) in the US currently infects between 2.0 and 3.9 million people; it is estimated that 80,000 will die annually from the infection, an average of 15 years before the general population (Mahajan et al., 2013), surpassing HIV as a cause of death for Americans (Ly et al., 2012). Many studies have attempted to estimate the origins of HCV as a human pathogen, but those dates are uncertain, ranging between 200 years and 150,000 years ago (Gonzalez-Perez et al., 1997; Smith et al., 1997; Simmonds, 2001, 2004). Although the HCV virus was only first described in the 1980s, Gray et al. (2013) were able to isolate HCV gene sequences from archived samples dating to 1953, identified in a previous study (Seeff et al., 2000) to date the origin of the US epidemic to 1901 (95% Cls: 1874–1926). Ancient genetics and phylogenetic methods inform our understanding of other historic outbreaks as well, not obfuscating the confirmation of Variola virus from mummified remains that date to a suspected smallpox pandemic in seventeenth and eighteenth century Siberia (Biagini et al., 2012). Most ancient virus research has been limited to historic or archival specimens, but a plant RNA virus sequenced from 750-year-old barley (Smith et al., 2014) illustrates the prospect of archaeological material as a source for human viral sequences in the near future. Estimating divergence times is also important for understanding the evolution of more slowly evolving pathogens. Leprosy, also known as Hansen’s disease, for example, is of interest to the public because of its persistence, clinical phenotype, and unique history of stigmatization. Mycobacterium leprae, the unculturable bacteria that causes the disease, was among the earliest pathogens targeted in skeletal remains because of its paleopathological signature in bone, with first reports of M. leprae PCR product published in the mid-1990s (Rafi et al., 1994). Since then many publications have targeted loci within M. leprae to identify the pathogen in archaeological remains (Haas et al., 2000; Montiel et al., 2003; Donoghue et al., 2005; Taylor et al., 2013), although suspiciously large PCR product sizes and/or inadequate authenticity measures casted doubts on initial reports, similar to many early PCR-exclusive aDNA studies (Stone et al., 2009). The recent technological advancements used to isolate, enrich, and sequence the Neanderthal genome (Burbano et al., 2010) and an ancient Yersinia pestis genome (Bos et al., 2011; Schuenemann et al., 2011) were applied to samples with signatures of leprosy. Schuenemann et al. (2013) and Mendum et al. (2014) used high-throughput sequencing, with and without array-capture, to reconstruct ancient M. leprae genomes from Medieval leprosy cases in Europe as well as several modern genomes, which had never been sequenced directly from a human infection. The ancient leprosy results were surprising. In fact, one sample displayed astonishing molecular preservation, with over 40% of reads mapping to M. leprae. There is a rare precedent for exceptional target DNA recovery (>70%; Meyer et al., 2012), but it is atypical for a pathogen derived from bone. Although mycobacterial DNA in particular has recently been shown to comprise 4–40% of overall microbial contaminant DNA in ancient horse samples (Der Sarkissian et al., 2014), the NGS leprosy data reflect a signature that is not consistent with carry-over of DNA from environmental contaminants, i.e., low average genome coverage with small regions of high coverage and diversity (Schuenemann et al., 2013). The authors attribute the unusually high proportion of M. leprae reads to the potentially decay-resistant effects of the mycotic acids.
unique to mycobacterial cell walls, also proposed by Zink et al. (2002) and Donoghue et al. (2004). Experimental verification of this hypothesis in mycobacteria other than *Mycobacterium smegmatis* would be worthwhile (Nguyen-Hieu et al., 2012).

Prior to the ancient leprosy genome data, scholars had questioned whether a change in virulence could explain the historically recognized decline of leprosy in the sixteenth century. The discovery of a close similarity between ancient and modern strains precludes a purely genetic cause for the changes in disease rates, emphasizing the importance of external factors in explaining such observations. Comparative genomic analyses between the ancient and modern strains of *M. leprae* reveal a remarkably reduced and conserved genome, with roughly 1,300 pseudogenes (Monot et al., 2009; Schuenemann et al., 2011). Using modern and ancient sequence data, tip sampling dates from 11 modern strains, and radiocarbon dates from archaeological samples, the authors calibrated substitution rates to infer that all human leprosy strains diverged at most ~5,000 years ago, consistent with paleopathological evidence but much earlier than previous estimations (Monot et al., 2005).

Still unanswered is the puzzling link between the clinical disease leprosy and another infectious agent, *Mycobacterium leprae* (subspecies *lepromatosis*). The closest known related species to *M. leprae* is *M. lepromatosis* and *M. bovis* (subspecies *bovis*). These species are unique to mycobacterial cell walls, also proposed by Zink et al. (2002) and Donoghue et al. (2004). Experimental verification of this hypothesis in mycobacteria other than *Mycobacterium smegmatis* would be worthwhile (Nguyen-Hieu et al., 2012).

A number of important and unexpected phylogenetic analyses have rejected the traditional ‘domestic animal-to-human’ hypothesis. The earliest and most notable case involves the discovery that human tuberculosis (TB) is ancestral to bovine TB, refuting the popular belief that bovids transmitted the disease to humans (Brosch et al., 2002). Likewise, phylogenies of taenid tapeworms show that humans were infected long before cattle and pig domesticates, probably as a result of scavenging during the Paleo-lithic, supporting a non-domestic origin of the parasite (De Queiroz and Alkire, 1998; Hobet al., 2000, 2001; Hobberg, 2006). Taenid tapeworms are one pathogen, like lice and *Helicobacter pylori*, that demonstrate a long coevolutionary relationship with modern and, in some cases, archaic humans and that are observed to have diversified and dispersed for millennia in parallel (Falush et al., 2003; Linz et al., 2007; Moody et al., 2012).

Did agriculture stimulate the emergence of the deadliest human diseases in history? To answer this important question we must understand disease ecologies that may influence the probability of a pathogen’s jump from another host to humans. Prehistoric communities did not occupy ‘pristine landscapes’ (Denevan, 1992). For example, if 10 million people occupied the Andes before European contact, as Denevan (1976, 1992) suggests, human impact on the environment, especially exploitation of natural resources, would have been extensive. Today, changes in land use and availability of resources as the result of human activity, in addition to climate change, lead to negative public health outcomes (Jones et al., 2013; Myers et al., 2013); principles of uniformitarianism would presuppose that the same mechanisms impacted health in the past, though likely at spatiotemporally-specific rates and/or scales, and with different consequences for humans and pathogens alike.

Each pathogen, however, shares a highly variable relationship with its host and thus benefits from individual examination. Pearce-Duvet (2006) reviewed several molecular phylogenetic analyses of pathogens for several human diseases associated with the rise of agriculture, including measles, pertussis, and falciparum malaria, and was unable to support or reject the domestic-origins hypothesis fully. This highlights the obstacles that define many of the research problems in pathogen evolution, mainly ambiguities caused by a limited number and diversity of available genetic data, the extinction of specific lineages, and biased signals due to demographic events (i.e., sweeps, bottlenecks). Technological improvements that make the field of modern and ancient genomics possible overcome such obstacles by increasing the amount of genetic data available for use in analyses and creating the potential to sample from extinct lineages. These data can then be used to date the divergence of animal and human forms of disease. These improvements may also enable us to look at how these pathogens acted as a selective force on humans at different time points in the past.

**Re-classifying the ‘heirloom’ and ‘souvenir’ pathogens**

In recent years, genetic evidence and phylogenetic reconstructions have been essential in testing hypotheses regarding the timing of pathogen emergence in human populations (see Pearce-Duvet, 2006), as discussed previously in regard to the ‘Paleolithic baseline.’ Pathogens thought to share a long coevolutionary history with humans or our hominin ancestors are referred to as “heirloom” species. Alternatively, ‘souvenir’ pathogens and parasites ostensibly were acquired after the NDT through the timing of pathogen emergence in human populations (see constructions have been essential in testing hypotheses regarding from where and what did pathological record? Discoveries with regard to these questions, in the clinical disease, especially as it is recognized in the paleo-mycobacterial cell walls, also proposed by Zink et al. (2002). The link between agriculture and the emergence of deadly pathogens prompted the traditional view that domestic animals transmitted diseases to their human caretakers (Weiss, 2001), who now had a population large enough to sustain them.

As it stands, genetic data, including complete genomes, are crucial to phylogenetic reconstructions, but they do not always resolve questions about the timing of a jump to humans or the timing of changes in virulence or patterns of transmission in an organism’s evolutionary history. The treponemal diseases, for example, remain enigmatic, as genetic data have not resolved the origins scenario for syphilis (see Mulligan et al., 2008; Harper and Armelagos, 2013). *Treponema pallidum*, the causative agent of one of which, venereal syphilis (subspecies *pallidum*), is sexually transmitted, whereas the others are not: yaws (subspecies *pertenuce*), bejel (subspecies *endemicum*), and pinta (subspecies *carateum*). The subspecies are morphologically
identical and for this reason, scientists have offered alternative biological, cultural, and climatic explanations for virulence (Hackett, 1963; Hudson, 1965). Yaws long has been thought to be an heirloom pathogen, infecting our ancestors as early as the Pleistocene (Rothschild et al., 1995) and also present in African apes (Harper et al., 2008). It remains unclear, however, whether syphilis arose in the Old World or in the New World and then was transmitted by returning explorers after contact (Hackett, 1963). This decades-long debate, reviewed in Baker and Armelagos (1988) and de Melo et al. (2010), has been revisited in recent years due to the availability of Treponema sequence data and the growing number of purported pre-Columbian paleopathological cases of skeletal treponematoses in the Old World (e.g., Cole and Waldron, 2011; Mays et al., 2012; Schwarz et al., 2013), although many of these cases have been refuted (Harper et al., 2011). Treponema, however, are uncultivable, which continues to pose problems for obtaining taxonomically and geographically broad sequence data for phylogenetic analyses. Numerous other factors complicate evolutionary analyses, including a lack of genetic diversity between subspecies, the similarity in skeletal lesions among the clinical diseases, the lack of comparative sequences, the eradication of extant species and strain diversity, and the difficulties isolating ancient Treponema DNA (Bouwman and Brown, 2005; Barnes and Thomas, 2006; von Hunnius et al., 2007).

The timing of syphilis’ emergence is also central to this discussion, but no studies have had the opportunity to use time-stamped sequence data or other methods of molecular calibration to test alternative hypotheses directly. Unfortunately, spirochetes are notoriously fragile bacteria, with unique structures that differ from their more robust GC-rich, gram-negative counterparts, such as Mycobacterium tuberculosis, rendering them even more susceptible to post-mortem degradation (Cox et al., 1992; Radolf and Lukehart, 2006). Furthermore, pathogen load is low during the disease stage where skeletal changes are observable, decreasing the likelihood of isolating Treponema DNA from archaeological remains. While ancient DNA sequence data theoretically would contribute to this debate, the unique traits of this bacterium generate pessimism. Bouwman and Brown (2005: 703) boldly but perhaps prematurely claim, “Ancient DNA cannot be used to study venereal syphilis” after finding no treponemal DNA in clear skeletal cases. Equally pessimistic are von Hunnius et al. (2007:2098), who concur with both Bouwman and Brown (2005) and Barnes and Thomas (2006) to send a warning: “…the discussion presented here and elsewhere, should serve as a warning for paleopathologists who hope that the new and exciting world of ancient DNA can help them with their observations of diseases in the past.” Often, the allure of the new methods, such as those used in ancient genetics, is associated with the tendency for scholars to forego grounded, problem-oriented research. In the case of Treponema, we should be hopeful that the methods and analytic tools will catch up to the questions the research community has sought to answer for decades. For example, Montiel et al. (2012) report T. pallidum pallidum in neotones dating from post-European exploration Spain, suggesting that active infection with congenital syphilis, especially in neonates, results in higher bacterial load and thus better spirochete DNA preservation and recovery. If this is true, new sampling strategies combined with enrichment techniques may yet be successful. Alternative methods of accessing preserved pathogen DNA, for example from dental calculus, may succeed as well.

Historically, we know that non-venereal treponematoses was found endemically worldwide. If treponemal disease diverged in Eurasia and traveled to the New World with human migrants, then pre-Columbian Old World Treponema genomes should harbor greater diversity, whereas New World strains would have experienced a bottleneck, as we see in other pathogens that evolved with a migrating host (e.g., Fisher et al., 2001). These are testable hypotheses if ancient Treponema DNA is successfully isolated and sequenced; the debate is thus ready for successes in ancient genomics.

Cholera, a free living souvenir?

Cholera, transmitted by contaminated water, is problematic for environments where adequate sanitation and clean water are lacking, and thus the disease is often associated with higher population densities. There have been seven Vibrio cholerae pandemics recorded since 1817; the most recent outbreak began in the early 1960s and continues to infect humans today, having been most recently brought to world attention in the 2010 Haiti epidemic (Colwell, 1996; Grad and Waldor, 2013). Although some scholars suggested the pathogen adapted suddenly to humans before its first detection in 1817 (McNicol and Doetsch, 1983), translated historical documents and accounts of travelers, reviewed in depth by Macpherson (1872), suggest the presence of a cholera-like disease predating the first documented human pandemic by hundreds or perhaps thousands of years (e.g., Bhishagaratna, 1963; Peters, 1885 [cited in Colwell, 1996]). Whether these early accounts describe V. cholerae or a different pathogen remains unknown. However, two genetically distinct 01 biotypes are thought to be responsible for the seven main human pandemics: classical and ‘El Tor.’ El Tor emerged during the sixth pandemic in the 1960s, and it was thus assumed the classical biotype was responsible for the pandemics prior to 1961. To test this hypothesis, Devault et al. (2014) reconstruct a V. cholerae genome from an ancient tissue sample archived during the second pandemic (1849) with targeted enrichment techniques and NGS, demonstrating that the second cholera pandemic was caused by a strain most-closely related to the classical V. cholerae biotype, as opposed to the El Tor type circulating today. Due to limitations of estimating divergence time using dates from only the terminal branches of the tree, the authors use a large El Tor dataset (Mutreja et al., 2011) to estimate subtitution rates. According to a strict molecular clock, the emergence of pathogenic V. cholerae occurred ~430–440 years ago; the authors, however, highlight the misleading effects of genetic recombination and saturation on dating estimates, concluding their data could also be consistent with a longer evolutionary history of human cholera, potentially coinciding with the first epidemiological transition during the NDT (Devault et al., 2014). Studies that integrate ancient samples into a large phylogenetic dataset inform our understanding of both past and recent epidemics. Interestingly, cholera is a free-living environmental organism and has also been used as an example of how human disease patterns respond to climate change (Colwell, 1996; Rodo et al., 2002). In the future, ancient samples may help elucidate how climactic events, like El Niño, played a role in ancient disease transmission to create awareness for prevention of potential outbreaks.

Phylogenies are not only strengthened by taxonomically-, geographically-, and temporally-representative data, but data representative of the entire genome. Combining genome scale data with the archaeological or historical contexts from which they are isolated enables examinations of how pathogens evolve or how conditions change to produce a fatal human outbreak.

Understanding the evolution of pathogens

What causes a pandemic? The timing of pathogen divergence events in phylogenetic reconstructions ultimately sheds light on the emergence of human disease throughout our history, but additional data are needed in order to infer the causative conditions, biological or environmental, that set the stage for human
pandemics. Biopsy or autopsy tissue specimens from patients during influenza outbreaks and skeletal remains from associated plague contexts provide an excellent source of data from which to reconstruct the evolutionary history of pandemics. The first ancient pathogen genome was recovered from the notorious 1918 ‘Spanish’ influenza pandemic (Taubenberger et al., 1997). Accomplished before the extensive use of NGS in aDNA research, this work represented a significant achievement and exemplifies the advantages of genome data for examining the evolutionary history and origins of disease. Outbreaks of Influenza A RNA viruses (genus Orthomyxoviridae) occur seasonally each year, but only occasionally result in extremely virulent pandemics. Most notable is the 1918 ‘Spanish’ influenza, which caused nearly 50 million deaths worldwide (Johnson and Mueller, 2002). In the late 1990s, Taubenberger et al. (1997) first isolated fragments of the RNA virus from formalin-fixed paraffin-embedded (FFPE) tissues taken from U.S. servicemen killed in the 1918 pandemic. More recent studies sequenced fragments of the virus from the frozen lung tissue of an Inuit woman buried in Alaskan permafrost since her death in 1918 (Reid and Taubenberger, 1999) and two additional fixed samples from 1918 influenza victims at the Royal London Hospital (Reid et al., 2003a).

The natural reservoir of Influenza A viruses is thought to be aquatic birds (Webster et al., 1992), and thus the phylogenetic relationships among avian, mammalian, and the 1918 human strains are critical to understanding the pandemic’s origins. There are eight RNA segments in the influenza virus genome, encoding at least 10 proteins. The virus phylogenies vary depending on the segments available for analysis, which increased as research progressed, from small gene fragments (Taubenberger et al., 1997), to full gene sequences (Reid et al., 1999, 2003b; Basler et al., 2001), and eventually all eight RNA segments—the entire genome (Taubenberger et al., 2005). For example, phylogenetic analyses of the haemagglutinin (HA) gene, a crucial element involved in host interaction, always place the 1918 strain with the mammalian rather than the avian clades, despite some similarities to HA in avian viruses (Reid et al., 1999). By contrast, nonsynonymous or amino acid changes observed in the 1918 neuraminidase segment result in a phylogenetic placement with the avian viruses (Reid et al., 2000). It is important to understand the different pathways from which a lethal pandemic can emerge. In this case, the evolutionary scenario neither resembles direct reassortment nor adaptation from an intermediate host (Reid and Taubenberger, 2003), but the sequences isolated from archived samples ultimately reflect an avian influenza-like genome, suggesting an avian common ancestor, with or without an intermediate host (Taubenberger et al., 2012).

The phylogenetic ambiguity observed with the Influenza A strains reflects broader issues crucial for the evolutionary analysis of pathogens: biases caused by each site/gene’s variable evolutionary history, insufficient taxonomic coverage and/or genetic diversity to resolve phylogenetic relationships, and an overall limited understanding and sampling of potential wild reservoirs/hosts. Ancient pathogen genome research must confront such challenges as it matches the speed of technological development. In 2013, a study applied NGS technology without enrichment to a previously reported archived tissue sample and produced in a single sequencing run a 3000x coverage genome of the 1918 Influenza A virus. Their analyses of this genome in comparison with that of the 2009 H1N1 pandemic virus found that the 1918 virus demonstrated significantly increased inflammatory and cell death responses (Xiao et al., 2013). This technical advance further demonstrates the promise of newly developing methods with ‘ancient’ medical specimens and highlights the excitement of obtaining DNA/RNA from temporally distinct outbreaks. It would be of great interest to apply these new methods to different samples to examine how the virus changed over the course of this pandemic.

Perhaps one of the most famous pandemics in history is the Black Death, one of three historical plague pandemics. The Black Death decimated the European population in the fourteenth century, and the twin papers reporting on a Yersinia pestis genome from fourteenth century plague victims are the first and clearest successes of targeted genome enrichment in ancient pathogen DNA (Bos et al., 2011; Schuenemann et al., 2011). Although many previous studies had reported the presence of Y. pestis in plague contexts (e.g., Raoult et al., 2000; Haensch et al., 2010; Wiechmann et al., 2010), Bos et al. (2011) and Schuenemann et al. (2011) ultimately had two clear advantages: a historically-documented Black Death cemetery with a tight temporal constraint (AD 1348–1350) and genome-scale sequence information. They confirmed Y. pestis as the causative agent of the Black Death, as opposed to a separate biotype as previously hypothesized (Devignat, 1951), and they concluded that its phylogenetic position placed the ancient genome as ancestral to all extant human strains (i.e., branch 1 and 2 isolates). These cemetery dates were used to calibrate the molecular clock, and their analyses traced the divergence time of all currently circulating isolates of branch 1 and 2 strains to the thirteenth century, proposed to correspond to an origin in or near China followed by plague transmission along the Silk Road (Morelli et al., 2012; Bos et al., 2013).

With a later publication, Cui et al. (2013) greatly expanded the number of modern Y. pestis whole genome sequences, which impacted the estimates for clock rate variation and thus challenged Bos et al.’s (2011) divergence dates. The authors hypothesized that transmission history affects substitution rates, such that epidemic periods result in more SNPs than enzootic phases, and that all Y. pestis strains are capable of causing disease in humans, not just branches 1 and 2. Furthermore, they emphasized that an analysis of fewer genomes or less geographically representative genomes would not have provided enough statistical power or resolution to demonstrate their observations from the full dataset. There were only 17 reference sequences available before Morelli et al.’s (2010) and Cui et al.’s (2013) collective contribution of over 300 new genomes. It is thus clear that including the full range of species diversity is important to phylogenetic resolution. Likewise, recent papers report genome sequences from the first ‘Justinian’ plague (AD 541–543), integrating these genomes with 131 Y. pestis strains from the following two pandemics to demonstrate that the earliest historic pandemic likely reflects an extinct outbreak, independent from the devastating Black Death pandemic 800 years later (Wagner et al., 2014). As with all paleogenetics studies, the ancient Yersinia publications were strengthened by the modern genome work and extensive sampling, and they highlight a weakness for studies of some current pathogens that are either unculturable (e.g., leprosy, syphilis), extinct, rare (e.g., those causing bejel, yaws, and smallpox), or neglected (e.g., those causing leishmaniasis, dengue, and schistosomiasis), in that often few comparative data are available.

The Black Death is also an important case for understanding the biological and demographic consequences of pandemics. DeWitte and Wood (2008) suggest, for example, that despite perceived virulence of the Black Death in London, the plague did not kill indiscriminately but reflected selective mortality with respect to frailty. In fact, no unique derived positions were found in the ancient genome to suggest a difference in virulence from currently circulating strains that would explain the massive death toll seen in the Black Death (Bos et al., 2011). As observed in work on HIV evolution (Worobey et al., 2010), recent research contradicts the outdated assumption that high virulence must be associated with a short-term relationship with a host (Read, 1994 [reviewed in Alizon et al., 2009]). If a genetic explanation is not plausible, perhaps evidence for an environmental one, such as adaptation to the vector, a
louse in the case of the Black Death (Ayyadurai et al., 2010), is yet to be uncovered. Virulence cannot always be explained clearly with genetics, and alternative factors, such as changes in vectors and variation in immune responses among individuals, as well as social conditions that affect response to disease and transmission (Armelagos et al., 2005), can also alter the impact of the same pathogen within and among populations.

The impact of ancient migration on disease transmission

Migration is one of the most influential processes impacting human-pathogen interactions: it facilitates the movement of existing human pathogens to new environments and/or communities, as well as the expansion of humans into new ecological niches, resulting in interactions with new vectors and disease reservoirs. Sequence data representative of the entire genome better facilitate inferences concerning biogeographical patterning of pathogens over time as they disperse or are carried across the landscape. In the increasingly globalized world, the sudden introduction of pathogens and parasites from a single traveler to naïve populations can have devastating effects, as witnessed in the recent 2010 Haiti cholera epidemic, which likely originated from Nepalese UN Security Forces during relief work (Chin et al., 2011; Hendriksen et al., 2011; Grad and Waldor, 2013). While globalization plays a role in current disease transmission, similar factors were important in the past, albeit at a slower pace.

The New World prior and subsequent to the Age of Exploration provides a unique context to explore the relationship between humans and infectious disease. As a result of European exploration and settlement, indigenous populations of the New World suffered massive demographic collapse, largely from the diseases transmitted during the ‘Columbian exchange’ (Crosby, 1972; Thornton, 1987; Verano and Ubelaker, 1992; Cook, 1998). Diseases present in the New World before European contact are inferred from skeletal or tissue changes consistent with current clinical manifestations. Unfortunately, the exact causative agents are often unknown, although recent ancient genome research is contributing to debates on New World disease origins.

Tuberculosis is one such disease that has been the focus of archaeological inquiry, likely owing to its conspicuous skeletal signature. Twentieth century physicians did not accept extra-pulmonary spinal lesions (Pott’s disease) as evidence of TB in the ancient New World (Hrdlicka, 1909; Morse, 1961; Cockburn, 1963) and thus proposed that TB was only present in the Old World prior to the Age of Exploration, concurring with the now outdated view that TB was acquired from cattle during domestication in the Old World (Rich, 1944; Cockburn, 1963). The high morbidity and mortality from TB in Native American populations following European contact also bolstered the post-contact Old World origin scenario, with New World groups considered ‘virgin soil’ for the organism. Because Native American groups reacted as if they were epidemiologically ‘naïve’ to the pathogen, for decades it was believed that Mycobacterium tuberculosis could not have existed in the ancient New World (Hrdlicka, 1909; Cockburn, 1963; Morse, 1961; Stead et al., 1995; Stead, 1997). There remains a lively debate regarding the disease’s New World origins, especially in light of new data from pre-Columbian mycobacterial genomes (Boš et al., 2014), discussed below.

Today, human strains of the M. tuberculosis complex (MTBC) are most diverse in Africa, while currently circulating strains in the Americas most closely resemble European lineages (Gagneux and Small, 2007), prompting the hypothesis that strains present in the New World before European exploration were either replaced or did not exist (Cockburn, 1963; Stone et al., 2009; Pepperell et al., 2013). Although decades of paleopathological evidence for pre-Columbian TB are now well accepted (see Roberts and Buikstra, 2003), until recently, molecular investigations using ancient pre-Columbian material, which identified short conserved regions of mobile elements considered to be diagnostic for TB, offered no information about phylogenetic placement, and are thus difficult to authenticate as ancient (e.g., Spigelman and Lemma, 1993; Salo et al., 1994; Arriaza et al., 1995; Shapiro and Gilbert, 2006; Klaus et al., 2010). The question thus remained: what precisely is New World TB?

Genomic-scale studies, capable of addressing the New World TB question, were first successful in the Old World with historic skeletal material from England and Hungary. The genotype of the nineteenth century TB strain found via DNA capture in remains buried at St. George’s Crypt in Yorkshire, England demonstrated a close genetic relationship with the modern reference strain H37Rv, not unexpected for the time and region (Boowman et al., 2012). In the Hungarian individual, Chan et al. (2013) use metagenomic sequencing to recover a mixed-strain infection, linking the strains to separate twentieth century European outbreaks. The authors claim that the 8% of TB reads are unlikely due to contamination because of the deep and even genomic coverage across H37Rv. Unfortunately, no damage profiles were generated to verify the patterns expected for ancient DNA; the sequence reads were, however, made publicly available. Muller et al. (2014) have reported two additional cases of multiple M. tuberculosis infection. Given the frequency of multiple strain infections in regions of high clinical TB risk today (Borgdorff and van Soolingen, 2013; Hingley-Wilson, 2013), mixed infection was likely common in the past. These discoveries in historic material highlight the question of whether or not genomic signatures of co-infection are recognizable in NGS data. In the case of the co-infected Hungarian mummy, a normal distribution of allele frequencies with a mean around 50% is shown to reflect the signature of two infections, and these could be disentangled computationally (Boš et al., 2014). Although ancient NGS library methods require that each sample receive its own barcode, the unintended transfer of barcodes between samples via ‘jumping PCR’ could mimic a signature of co-infection, appearing as sequence reads from more than one source (Kircher et al., 2011b). It is unclear whether multiple infections will be discriminable, an obvious obstacle for accurate evolutionary and phylogenetic inferences. Future research will likely require investigations of co-infected clinical TB patients combined with the development of new computational tools.

In spite of the challenges, the studies of Old World TB can be cited as proof of principle that TB DNA preservation can be robust enough for evolutionary analyses and thus similar methods were applied to New World samples. Following the application of in-solution and on-array capture to screen for and target the entire MTBC genome, Boš et al. (2014) confirm that a member of the MTBC caused infection in three pre-Columbian humans, all associated with the Chiribaya cultural tradition in southern Peru. Additionally, DNA damage profiles displayed patterns expected of ancient molecules, helping to rule out false-positive exogenous mycobacterial sequences. Surprisingly, the Peruvian strains are most closely related to TB adapted to seals and sea lions, distinct from all known human-adapted forms. This result rejects the hypothesis that New World TB strains are most closely related to strains in Asia, as would be expected if they arrived in the New World over the Bering Land Bridge with humans. The archaeological evidence for persistent and heavy reliance on seals and sea lions along the coast for millennia provides a plausible explanation for the transmission of Mycobacterium pinnipedi to humans. The lack of any ancient New World genomes from outside of the Osmore River valley in Peru precludes, however, confirmation of whether this strain was transmissible among humans or whether or not it is responsible for...
the abundance of skeletal cases that culminate towards the end of the first millennium AD in both South and North America (Roberts and Buikstra, 2003).

More surprising and potentially contentious are the results from dating analyses using the ancient TB genome data to calibrate divergence dates (Bos et al., 2014), which contradict the most recent modern genomic analyses (Gutierrez et al., 2005; Comas et al., 2013) and long-held views on the antiquity of tuberculosis. Whether the MTBC is the result of a zoonotic jump within the last few thousand years or an heirloom pathogen, emerging in Africa and dispersing with human migration, has been one of most enduring questions in pathogen evolution. Using models of demographic history to generate coalescent dates, Comas et al. (2013) concluded that TB likely dispersed out of Africa with human hosts at the end of the Pleistocene. The Peruvian genomes, whose radiocarbon dates effectively function only as a single calibration point, and the Hungarian individual of known death in 1797 are used independently to suggest a most recent common ancestor (MRCA) for the MTBC less than 6,000 years ago (Bos et al., 2014). The poorly understood effects of horizontal gene transfer of degraded DNA (Overballe-Petersen et al., 2013) and rate heterogeneity could contribute to inaccuracies in the dating methods. However, the divergence times are in accordance with other recent work (Namouchi et al., 2012), including modern epidemiological data (Bryant et al., 2013). Reports identifying tuberculosis via macroscopic and molecular methods in human remains that predate the proposed MRCA of the TB complex (Rothschild et al., 2001; Hershkovitz et al., 2008) have received strong criticism for unconvincing paleopathological and molecular evidence (Stone et al., 2009; Wilbur et al., 2009). If possible, applying newer NGS methods, including enrichment strategies, would allow for more robust measures of authenticity to early samples, providing more insight into the genetic relationship between tuberculosis-like infections in humans and currently known mycobacteria.

Paleogenomics has its limitations

Genomics is not a panacea, of course. Genomes are not always capable of resolving evolutionary relationships. Researchers also must contend with massive computational loads and face the immense task of sequence assembly, among other challenges. In this way, ancient genomics faces the same obstacles as modern genomics but is worsened by the inherent problems of isolating degraded and damaged DNA. Aside from the profound methodological and computational challenges, there are limits to aDNA analyses, especially when working with skeletal remains. While diseases such as tuberculosis, leprosy, and syphilis have paleopathological signatures that develop over time, many diseases do not manifest skeletally, or are fatal before they can. Where evidence exists for a pathogen in human remains, it is difficult to infer morbidity of that individual since much remains unknown about the pathogenicity (i.e., does the organism cause disease?) and/or virulence of many microbes. In addition, not all pathogens present in the past exist today, and the species that most commonly cause disease in humans today may not be exclusively responsible for disease in the past.

By solely focusing on skeletal remains to identify disease, we are limited to a small portion of the disease load in any population (Roberts and Manchester, 2005). This is partially due to a phenomenon Wood et al. (1992) described as the osteological paradox: in the burial record, those who appear healthy in the burial record were the sickest. Conversely, individuals with chronic signs of pathology were ‘healthier’ in life—immunologically capable of fighting infection long enough for the formation of skeletal lesions. Reluctance to use destructive analyses on skeletal remains without pathological lesions is justifiable but does restrict sampling to those with visible lesions, or those interred in cemeteries historically associated with plagues or epidemics (e.g., Baziotopoulou-Valavani, 2002; Schuenemann et al., 2011; Harbeck et al., 2013). A potential solution, however, is new methodology that requires very little bone or tooth powder (<50 mg) that may mitigate some concerns regarding sample destruction.

Other caveats inherent to studying pathogens in archaeological remains, such as poor burial preservation, uncertain differential diagnoses, and differential burial treatment of the sick, further limit what samples are available for ancient DNA analyses. Finally, there is the reality that some bacterial DNA and most virus RNA/DNA are not stable over time and may never be isolated from ancient remains (Bouwman and Brown, 2005; Barnes and Thomas, 2006), regardless of the technological advancements. As mentioned here and examined thoroughly elsewhere (e.g., Mitchell et al., 2005; Briggs et al., 2007; Gilbert et al., 2007; Dabney et al., 2013b), ancient DNA is subject to degradation, chemical alteration, and contamination, which can complicate or sometimes prevent extraction, sequencing, and downstream analysis. In the face of these obstacles, information from other biomolecules, e.g., mycolic acids and proteins, can complement ancient genomic studies. Tran et al. (2011) review the alternative methodologies and the resulting advances seen in ancient pathogen research.

One such promising alternative is dental calculus. Dental calculus has unique properties that may propel the field forward by enabling the isolation of an individual’s microbiome (as well as his or her genome) in cases where bone and tissue fail to preserve biomolecules (Preus et al., 2011). The unique human microbiomes reflect coevolutionary pressures that act both on the level of the host and on microbiota, whether mutualistic or pathogenic. Warinner et al. (2014) clearly demonstrate this observation and are able to use the oral microbiome to link diet, pathogens, and host immunity, a fascinating combination. In light of examining assumptions associated with the health of pre- and post-agricultural groups, we look forward to understanding whether the pre- and post-Neolithic microbiome differ; some of this exciting work has already begun (Adler et al., 2013).

Finally, there remain technological limits to recovering and reconstructing ancient microbial genomes, many of which are only surfacing as the field progresses. It is the case that experimental design is constrained by the expectation of what pathogen to target and what genomes are available for mapping sequence data. In rare cases NGS without enrichment and de novo assembly could surmount these issues but are currently exceptions to the standard. Some scholars prefer to use an unbiased metagenomic shotgun sequencing method approach to seek out pathogen DNA rather than targeted enrichment of one or a few species (e.g., Kay et al., 2014; Maixner et al., 2014). There is a fine line, however, between casting a wider net to take advantage of the strengths of NGS and conducting research haphazardly. In a press release by the American Society for Microbiology for a recent ancient pathogen study, a researcher stated, “We’re cranking through all of these samples, and we’re hopeful that we’re going to find new things” (Siwa, 2014). The risk of this approach is a study that lacks well-reasoned hypothesis testing and ultimately fails to make significant contributions to broader evolutionary questions because of an absence of appropriate problem orientation. The ‘let’s see what we might find’ attitude will not work well in all ancient DNA contexts.

Future directions

Ancient pathogen research has focused heavily on methodological challenges: isolating degraded molecules that are present in miniscule proportions. As technological sophistication increases,
however, other unresolved issues become apparent. For example, to interpret the evolutionary history of a pathogen we must explicitly consider the evolution of the pathogen in terms of co-evolution with the human host. Infectious disease, especially at pandemic levels, has the power to alter population demographics and affect cultural systems and modes of behavior. As discovered with the transition to farming, these changes are often accompanied by complex consequences for human health, many of which are currently being examined.

Ancient genomics is mutually beneficial to research in modern genetics, as both fields strive to identify polymorphisms that may have been important for human adaptation to pathogens or vice-versa over time. Both modern and ancient human evolutionary genomics now enable the detection of changes unique to human lineages by comparing human genomes with the available Neanderthal and Denisovan genomes to seek out changes linked to immune response and/or susceptibility (see review in O’Blenna et al., 2012). We also foresee the potential for population-scale surveys of ancient human exomes and genomes across sites or skeletal series in the future, especially with respect to identifying the selective effects of infectious disease over time and particularly in pre/post-pandemic eras, or as a result of the transition to agriculture. On a small scale, the nascent stages of this research began a number of years ago, most notably with the investigation of the CCR5 delta-32 allele frequencies in ancient European skeletal collections (e.g., Wang et al., 2012). The allele, which confers HIV-1 resistance, is thought to have arisen during a single event between 700 and 3,500 years ago, possibly coinciding with selection pressures in Europe following the Black Death or smallpox epidemics (Martinson et al., 1997; Libert et al., 1998; Galvani and Slatkin, 2003). This allele was then detected in comparable frequencies in Bronze Age skeletal remains and in post-plague remains in Central Europe, demonstrating that the Black Death was not a selective force influencing the increase in CCR5 delta-32 frequencies observed in northeastern European populations (Hummel et al., 2005). More recently, Olalde et al. (2014) take a more sophisticated approach to detecting this allele and other polymorphisms that may have influenced susceptibility to infection in European populations. The authors report on a newly sequenced Mesolithic genome with which they examine the genotypes at a number of loci implicated in immune response and host resistance. The early date of this human genome allows the authors to directly examine hypotheses about the remodeling of the immune system purportedly due to post-Neolithic events in Europe; they find a mixed picture, however, with some variants, thought to be derived, as ancestral, while others were confirmed absent in the Mesolithic individual (Olalde et al., 2014).

Recent paleogenomic research has not only questioned the antiquity of human adaptation to disease but also challenged the antiquity of traits thought to be adaptations on the part of the antiquity of human adaptation to disease but also challenged the antiquity of traits thought to be adaptations on the part of the human host. Infectious disease, especially at pandemic levels, has the power to alter population demographics and affect cultural systems and modes of behavior. As discovered with the transition to farming, these changes are often accompanied by complex consequences for human health, many of which are currently being examined.

Recent paleogenomic research has not only questioned the antiquity of human adaptation to disease but also challenged the antiquity of traits thought to be adaptations on the part of the human host. Infectious disease, especially at pandemic levels, has the power to alter population demographics and affect cultural systems and modes of behavior. As discovered with the transition to farming, these changes are often accompanied by complex consequences for human health, many of which are currently being examined.

Ancient genomics is mutually beneficial to research in modern genetics, as both fields strive to identify polymorphisms that may have been important for human adaptation to pathogens or vice-versa over time. Both modern and ancient human evolutionary genomics now enable the detection of changes unique to human lineages by comparing human genomes with the available Neanderthal and Denisovan genomes to seek out changes linked to immune response and/or susceptibility (see review in O’Blenna et al., 2012). We also foresee the potential for population-scale surveys of ancient human exomes and genomes across sites or skeletal series in the future, especially with respect to identifying the selective effects of infectious disease over time and particularly in pre/post-pandemic eras, or as a result of the transition to agriculture. On a small scale, the nascent stages of this research began a number of years ago, most notably with the investigation of the CCR5 delta-32 allele frequencies in ancient European skeletal collections (e.g., Wang et al., 2012). The allele, which confers HIV-1 resistance, is thought to have arisen during a single event between 700 and 3,500 years ago, possibly coinciding with selection pressures in Europe following the Black Death or smallpox epidemics (Martinson et al., 1997; Libert et al., 1998; Galvani and Slatkin, 2003). This allele was then detected in comparable frequencies in Bronze Age skeletal remains and in post-plague remains in Central Europe, demonstrating that the Black Death was not a selective force influencing the increase in CCR5 delta-32 frequencies observed in northeastern European populations (Hummel et al., 2005). More recently, Olalde et al. (2014) take a more sophisticated approach to detecting this allele and other polymorphisms that may have influenced susceptibility to infection in European populations. The authors report on a newly sequenced Mesolithic genome with which they examine the genotypes at a number of loci implicated in immune response and host resistance. The early date of this human genome allows the authors to directly examine hypotheses about the remodeling of the immune system purportedly due to post-Neolithic events in Europe; they find a mixed picture, however, with some variants, thought to be derived, as ancestral, while others were confirmed absent in the Mesolithic individual (Olalde et al., 2014).

Recent paleogenomic research has not only questioned the antiquity of human adaptation to disease but also challenged the antiquity of traits thought to be adaptations on the part of the pathogen. For example, D’Costa et al. (2011) identify loci implicated in antibiotic resistance in the metagenome of 30,000-year-old Beringian permafrost sediments, refuting the notion that these alleles are an exclusively modern phenomenon. Future ancient genome and metagenome reconstructions from various eras and archaeological contexts are imminent; eventually we will see a more fine-grained picture of circulating genetic diversity at numerous points in history, and conceivably the variability of individual responses to infection within a population. These population-level inferences will be facilitated by the increasing adoption of NGS methods in ancient DNA laboratories worldwide. Finally, one of the main benefits of ancient genomics adopting NGS methods is the potential to immortalize precious material. These ancient DNA sequencing libraries are becoming standard, such that over time, the thousands of skeletal and mummified remains processed by labs worldwide will generate a new fount of samples to target the host or any number of organisms within; a single aDNA library theoretically could contribute to a multitude of interesting research questions, and could better justify less focused metagenomic approaches. As costs decrease and techniques improve, including new methods that seek only to include damaged molecules during library preparation (Gansauge and Meyer, 2014), one can imagine archives of accessible ancient DNA libraries, removing the need for future destructive analyses, and forging the way for future collaborations and research in ancient population genomics.

**Conclusion**

It has been asked, “does genomic evidence raise more questions than it answers?” (Harper and Armelagos, 2013: 137). The answer is obviously, yes. Every infectious disease emerges within a unique spatiotemporal context. Migration, agriculture (including animal domestication), anthropogenic change to the landscape, shifts in cultural, behavioral, or institutional practices, and, most recently, increasing globalization and wide-reaching human interaction have all influenced our coevolutionary history with pathogens. Although recent genomics research has called into question exactly which diseases were introduced before or after the development of agriculture and domestication (see review in Harper and Armelagos, 2013), these practices indisputably had an epidemiological impact on human societies and on human biology. Archaeologists and biologists continue to debate whether the epidemiological impact resulted in increased mortality (Cohen and Armelagos, 1984; Steckel and Rose, 2002) or overall healthier populations (Gage and Dewitte, 2009; Pinhasi and Stock, 2011). Olalde et al. (2014) review genome-wide association studies that address the susceptibility of human populations to disease, revealing the complex relationship of background genetics on disease expression that is often population specific. The microbiome, representing the particular suite of organisms existing and coevolving in a single human host, helps shape the individual genetic background that influences a host’s immune response. Soon the field will combine information about pathogen evolution on a particular host background and on multiple scales, i.e., community and individual, to draw inferences about human health in the past, and in the future.

Methodological improvements for the isolation, sequencing, and authentication of ancient genomes will continue to result in a number of important contributions to the field. Genomic-scale sequence information from historical, archaeological, or paleoanthropological timescales can better address both macro and micro-evolutionary questions, whether sourced from humans or microbes. Nevertheless, taxonomic or biogeographic coverage is as crucial as genomic coverage in accurate phylogenetic inferences, urging more comprehensive sampling strategies. Phylogenetic inference, dating calibration, disease origins, host–pathogen interactions and coevolution, biogeographical patterns, including local extinctions, ancient population structure, and paleopathological diagnoses are a few of the research areas or methods that will benefit from continuing work in both ancient and modern pathogen genomics. Understanding the evolution of human infectious disease has as much practical importance as it does scholarly interest, especially in fields of medicine as we enter the era of personalized genomics. Until the next big revolution in the field of ancient DNA, enrichment, NGS, and phylogenomics are an exciting way forward in ancient pathogen research.

**Acknowledgments**

We would like to acknowledge the following sources of funding: The National Science Foundation for NSF BCS-1063939 and BCS-


