Journal of Human Evolution 79 (2015) 137–149

Contents lists available at ScienceDirect

Journal of Human Evolution

journal homepage: www.elsevier.com/locate/jhevol

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

ARTICLE INFO

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

Keywords:
Ancient DNA
Pathogen evolution
Human disease

ABSTRACT

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.

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 socio-cultural 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 NGS 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

* Corresponding author.
E-mail address: kmharkin@ucsc.edu (K.M. Harkins).
[1] Current address: Department of Anthropology, University of California, Santa Cruz, CA, United States

http://dx.doi.org/10.1016/j.jhevol.2014.11.002
0047-2484/

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 phylogenetic 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 (Bos et al., 2011; Schuenemann et al., 2011; Achtman, 2012), leprosy (Schuenemann et al., 2013), tuberculosis (Bouwman et al., 2012; Chan et al., 2013; Bos 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 demography 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). Epidemiological 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., 1998) 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; Debruyne and Poinar, 2009; Ho and Phillips, 2009). Sample-dating errors, however, appear to inflict 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.