

**Insights into microbial evolution and ecology
from genetic analysis of diverse archaeological
materials**

Dissertation

To Fulfill the
Requirements for the Degree of
„Doctor of Philosophy“(PhD)

Submitted to the Council of the Faculty
of Biological Sciences
of the Friedrich Schiller University Jena

by M.Sc. Archaeological Sciences Susanna Sabin
born on 06 March 1995 in New York City

Gutachter:

1. Prof. Dr. Johannes Krause (Max Planck Institute for the Science of Human History, Jena)
2. Prof. Dr. Christina Warinner (Max Planck Institute for the Science of Human History, Jena)
3. Dr. Iñaki Comas (Biomedicine Institute of Valencia, Valencia [Spain])

Beginn der Promotion: 19 Januar 2017

Dissertation eingereicht am: 24 Juni 2019

Tag der öffentlichen Verteidigung: 01 November 2019

Table of Contents

1. Introduction.....	4
1.1 Ancient microbial genomics.....	5
1.2 Shifting narratives on the origins of tuberculosis.....	8
1.2.1 The rise and fall of the Mycobacterium bovis hypothesis.....	8
1.2.2 Ancient DNA and tuberculosis.....	9
1.2.3 Discrepancies between different lines of evidence.....	10
1.3 Ancient metagenomics of human microbiota.....	11
1.3.1 Dental calculus.....	12
1.3.2 Paleofeces and coprolites.....	13
2. Aim of the thesis.....	14
3. Overview of manuscripts and author's contribution.....	16
3.1 Manuscript A.....	16
3.2 Manuscript B.....	17
3.3 Manuscript C.....	18
4. Manuscript A.....	20
5. Manuscript B.....	76
6. Manuscript C.....	102
7. Discussion.....	132
7.1 Challenges in uncovering the origins of tuberculosis.....	132
7.1.1 Archaeological sources of MTBC DNA.....	132
7.1.2 Authentication and environmental mycobacteria.....	134
7.1.3 Challenges of demographic inference and molecular dating using ancient DNA	134
7.1.4 Where to aim for future studies.....	136
7.2 Characterizing ancient human microbiota.....	137
7.2.1 Facilitating qualitative interpretability of large metagenomic datasets.....	137
7.2.2 Limitations of modern databases and taxonomic classifiers.....	138
7.2.3 Inferring microbiome dynamics and "normal" microbiomes.....	138
7.2.4 Dentin as a potential source for oral microbes.....	139
7.3 Challenges in studying ancient eukaryotic parasites with ancient DNA.....	140
7.4 Conclusion.....	141
8. References.....	143
9. Summary.....	162
10. Zusammenfassung.....	164
11. Eigenständigkeitserklärung.....	166
12. Curriculum Vitae.....	167
13. Acknowledgments.....	171

1. Introduction

The study of human health over time offers valuable pathways for understanding multiple aspects of human experience and biology. Determining the presence of a disease in an ancient individual or community can give us insights into daily life during that time, and comparing human microbiota between different human groups over time and space can offer insights into behavior and diet. Assessing the health of past populations may provide new perspectives on concomitant social or political changes, and contribute to our understanding of how those populations managed, or failed to manage, crises and change. On a broader level, identifying and interrogating humanity's relationship with infectious and commensal microbes may offer insights into human evolution and adaptation. Most hopefully, the knowledge gained from the basic science of past human health may one day lead to discoveries that can be applied to modern medicine. For example, the evolutionary history of a specific pathogen may allow us to understand how it may behave in the future, and the constitution of ancient human microbiota may allow us to interrogate what taxa have been gained and lost over time in certain populations and what this may mean for modern oral and gut health.

The study of past human health has always, by necessity, been an interdisciplinary endeavor. The task of diagnosis, difficult in living populations, becomes increasingly complicated with the passage of time, and the meaning and value of historical diagnosis, depending on the theoretical tides among medical historians, modern clinicians, and anthropologists, may fluctuate (Arrizabalaga, 2002; Waldron, 2009). Historical documentation or art pieces may offer verbal descriptions or visual depictions of ill health, but may be open to broad interpretation due to differing medical conventions and terminology over time and space, embellishment of the artist, or even political concerns that may or may not be evident to the scholar attempting a diagnosis (Mitchell, 2011). Health can also be inferred from human remains in archaeological contexts or the archaeological contexts themselves. Mass graves or multiple burials, for example, could signal an epidemic event (Blakely and Detweiler-Blakely, 1989; Rugg, 2000). Disease processes can leave traces in surviving soft tissue in the case of mummified individuals or in hard tissue, and as human remains that survive in the archaeological record are mostly skeletonized, bone tends to be the most common medium from which paleopathologists attempt to diagnose deceased, archaeological individuals. However, pathological changes in bone are not always specific to a single condition or infection, and not all conditions and infections leave signs in the

skeleton. For those conditions that *can* leave signs in the skeleton, whether or not this occurs depends on numerous factors, such as, for example, the immunocompetency of the individual or the severity of the infection. Even in the event that there are pathognomonic signs of a specific infection identified in a skeletonized individual, that information cannot be used to infer evolutionary dynamics of the infecting organism.

Archaeogenetics, with existing technologies and in its current state-of-the-art, can offer direct molecular evidence of individual microbes and microbial communities in the past. If we are able to pass the threshold of acquiring trustworthy ancient DNA, we can tell the story of evolutionary change in a pathogen down to the level of a single mutation, and illustrate the similarities and differences between human-associated communities of microorganisms over time and space. However, microbial archaeogenetics is still a young field. This thesis illustrates the potential of ancient DNA to answer relevant questions about human health by zooming in to a single pathogen genome enriched from a metagenomic background and zooming out to whole microbial communities extracted from archaeological materials. Here, the origins of a family of pathogens – the *Mycobacterium tuberculosis* complex (MTBC) – is further elucidated with a single ancient genome, and metagenomics is leveraged to explore preservation dynamics for human-associated microbial DNA in different archaeological contexts. The enclosed projects emerged from metagenomic shotgun screening of DNA extracted from a calcified lung nodule, dentin, dental calculus, and bulk latrine sediments. The work presented here builds upon decades of developments in genetics, archaeogenetics, and microbiology. As an introduction, I will focus on the major developments that have specifically led to the state of the art in ancient microbial genomics, then go into greater detail describing the history of scholarship investigating the origins of tuberculosis and archaeological metagenomics of human-associated tissues.

1.1 Ancient microbial genomics

The first ancient DNA studies were published in 1984, and featured the cloning of DNA fragments from an ancient Egyptian mummy (Pääbo, 1984) and the extinct Quagga, or plains zebra (Higuchi et al., 1984). In the following years, the polymerase chain reaction (PCR) method of DNA duplication was invented, and the door to studying ancient DNA from pathogens was opened (Mullis and Faloona, 1987; Pääbo, 1989). The first study presenting genetic evidence of an ancient microbe from archaeological human bone was released in 1993, and reported the amplification of *Mycobacterium tuberculosis* complex (MTBC) DNA (Spigelman and Lemma, 1993). Shortly thereafter, Salo and colleagues amplified MTBC DNA from mummified human tissue (Salo et al., 1994). The amplification of *Yersinia pestis* from plague victims followed, launching the practice of targeting the tooth pulp chamber for evidence of septicemic pathogens in human remains (Drancourt et al., 1998). Though these

studies were innovative, they triggered skepticism, and guidelines and methodologies for ensuring the veracity of results stemming from ancient DNA were required. To limit the possibility of cross-over contamination from modern or other ancient *Y. pestis* studies, Raoult and colleagues presented a methodology they termed “suicide PCR,” in which unique, never-before-used primers were used to amplify ancient *Y. pestis* DNA from archaeological samples (Raoult et al., 2000). The experiments would be performed without a positive control to further avoid contamination. However, this method was accompanied by many practical limitations. Different primer designs with overlapping regions could not be tested and refined in-house. It limited within-lab reproducibility and scalability, as according to the suicide PCR protocol, the selected primers could only be used once within a given laboratory, lest the resulting PCR product contaminate future experiments. In a sense, reproducibility between labs was also limited in the case of a single sample, because whatever remained of the sample would in theory be contaminated with its own product from a given primer set. Reproducibility became a central aspect of ancient DNA studies as the field grappled with how to authenticate findings (Richards et al., 1995; Roberts and Ingham, 2008; Wilbur et al., 2009). It became clear that the standards for ancient DNA authentication would vary depending on the type of data targeted and the study design. Gilbert and colleagues outlined a flexible, “cognitive” framework for authentication taking into consideration the risk of contamination for various types of target organisms (Gilbert et al., 2005). Ancient pathogen DNA was cited as the second-highest risk category of study target in archaeogenetics in terms of contamination, just below human DNA (Gilbert et al., 2005).

Next generation sequencing (NGS) transformed the field of archaeogenetics, allowing data to be produced indiscriminately at high volumes (Mardis, 2008). The first aDNA studies leveraging high-throughput sequencing presented portions of nuclear DNA from mammoths (Poinar et al., 2006) and Neandertals (Green et al., 2006). The variability of environmental and microbial contamination across different preservation environments was demonstrated within these first studies, with the mammoth analysis yielding ~45% endogenous DNA against a heavy microbial background (Poinar et al., 2006). NGS also allowed for the consistent use of the chemical degradation of DNA over time (Lindahl, 1993) as a tool for authentication. Using 454-sequenced Neandertal DNA, Briggs and colleagues confirmed the overabundance of purines at fragment breaks that had previously been reported in PCR product (Hofreiter et al., 2001) and postulated that this was a reliable pattern for ancient DNA (2007) that could not be observed to the same degree in more recent DNA (Sawyer et al., 2012) or present day contamination (Krause et al., 2010). Since then, rate and placement of cytosine misincorporation along DNA strands have been reported as evidence for the antiquity of genetic material, and tools quantifying this phenomenon are now a standard part of aDNA analysis pipelines (Ginolhac et al., 2011; Neukamm and Peltzer, 2018).

Following the emergence of NGS, the reconstruction of ancient pathogen genomes became plausible through shotgun sequencing or capture (Burbano et al., 2010; Hodges et al., 2009; Maricic et al., 2010). The first ancient pathogen genome was generated using an array-based capture technique (Bos et al., 2011). The study, based on DNA extracted from dental pulp chambers of individuals buried in a “plague pit” in London, further confirmed *Y. pestis* to be the most likely cause of the Black Death, and showed there to be little difference between the 14th century bacterial genome and its modern counterparts (Bos et al., 2011). This would be the first in many publications from archaeogenetics that would change how we understood the plague and its relationship to human history and movement – a story that is still evolving (Andrades Valtueña et al., 2017; Bos et al., 2016; Feldman et al., 2016, 2016; Keller et al., 2018; Rascovan et al., 2019; Rasmussen et al., 2015; Spyrou et al., 2018b, 2018a; Wagner et al., 2014). Subsequently, capture techniques were used to increase genomic coverage of microbes such as ancient *Mycobacterium pinnipedii* from vertebrae (Bos et al., 2014), *Mycobacterium leprae* from bones and teeth (Schuenemann et al., 2013), *Salmonella enterica* from teeth (Vågene et al., 2018; Zhou et al., 2018), *Vibrio cholerae* from preserved intestine (Devault et al., 2014), *Helicobacter pylori* from mummified gastrointestinal tissue (Maixner et al., 2016), and Hepatitis B virus from bone and mummified tissue (Mühlemann et al., 2018; Ross et al., 2018). In addition, researchers were able to acquire genome-level coverage from non-targeted shotgun sequencing alone for pathogens including *Mycobacterium tuberculosis* from mummified tissue (Chan et al., 2013; Kay et al., 2015); *Brucella melitensis* from a calcified nodule (Kay et al., 2014); *Gardnerella vaginalis* and *Staphylococcus saprophyticus* from calcified nodules (Devault et al., 2017); *Borrelia recurrentis* from teeth (Guellil et al., 2018); and Hepatitis B virus from teeth (Krause-Kyora et al., 2018).

In addition to bacteria and viruses, ancient eukaryotic pathogens have been detected in data produced with NGS technologies, though there have been no genomic reconstructions to date. DNA from *Plasmodium* spp., the causative agents of malaria, was isolated from human remains buried in southern Italy during the Imperial Roman period (Marciniak et al., 2016) and blood slides from a mid-twentieth century medical collection (Gelabert et al., 2016). Genetic material from human intestinal parasites, such as *Ascaris lumbricoides* and *Taenia* spp., has also been sequenced from coprolites and latrine sediments (Côté et al., 2016; Sørensen et al., 2018). As discussed in this thesis, eukaryotic parasites are thought to play an important role in human evolution, and can act as proxies for human diet and migration (Mitchell, 2013; Perry, 2014). However, there is currently a dearth of curated genetic sequences of eukaryotic parasites for archaeogeneticists to reference, though there are active research projects aiming to remedy this data gap (Aurrecochea et al., 2013; International Helminth Genomes Consortium, 2019).

Such studies have been instrumental in elucidating the history of human health in ways that would have been impossible with modern genetics alone. The power of aDNA to transform the narrative of a disease's history is discussed further below, with the specific case of tuberculosis.

1.2 Shifting narratives on the origins of tuberculosis

Tuberculosis is an infectious disease that is estimated to infect approximately one quarter of the global human population (Houben and Dodd, 2016). The family of bacteria responsible for causing tuberculosis is the MTBC, which is made up of seven human-associated lineages and one cluster of animal-associated lineages. Manuscript A of this thesis presents evidence supporting the hypothesis that the extant members of the MTBC, excluding the smooth tubercle bacillus *Mycobacterium canettii*, emerged no more than 6,000 years before present. The story of the MTBC has oscillated between either one of an ancient family of pathogenic bacteria that precede the emergence of agriculture or a group of bacteria that emerged as human pathogens for the first time following the advent of cattle domestication.

1.2.1 *The rise and fall of the Mycobacterium bovis hypothesis*

It was long hypothesized that tuberculosis and similar infectious diseases could not maintain themselves in human populations prior to the rise of agriculture and sedentary, urbanized living (Armelagos et al., 2005; Dobson and Carper, 1996; Pearce-Duvel, 2006). The existence of *Mycobacterium bovis*, a species of the *Mycobacterium tuberculosis* complex (MTBC) that predominantly infects domestic bovids, is closely related to the strains from the MTBC that routinely infected humans. Its capability of transferring from bovids to humans through tainted milk or meat products originally seemed to suggest that tuberculosis in human populations had originated as cross-over event of *M. bovis* (Pearce-Duvel, 2006). However, in the genomics era, it became clear based on patterns of gene loss in the MTBC that *M. bovis* and other animal-infecting strains constituted a more derived lineage than human-associated strains of tuberculosis (Brosch et al., 2002; Gagneux, 2018; Gutacker et al., 2002; Mostowy et al., 2002). Following this revelation, tuberculosis became conceptualized as an ancient disease, perhaps originating in human ancestors in concurrence with or prior to the emergence of "modern" *Homo sapiens* (Brosch et al., 2002; Gutierrez et al., 2005). This was supported by the apparent phylogeographic structure of the MTBC, by which it was inferred by some researchers that the different lineages of the MTBC were adapted to different human populations (Gagneux et al., 2006; Hershberg et al., 2008). However, these estimates were based on what were thought to be universal synonymous substitution rates across all bacteria (Gutierrez et al., 2005). Thus, they were also based on

the assumption that all bacteria evolved at the same rate, and their molecular clocks operated with the same speed. After it was demonstrated that bacteria had diverse substitution rates between and within taxa (Achtman, 2016; Cui et al., 2013; Ochman et al., 1999), it became necessary to evaluate the clock-like behavior of the MTBC given possible variability in substitution rates over time throughout the complex, and calibrate such analyses.

Prior to 2014, there was no ancient genomic data for the MTBC with which to directly calibrate the phylogeny. However, it had been shown across multiple taxa that molecular clocks were dependent on the window of sampling times, with younger calibration points being associated with inflated substitution rates (Achtman, 2016; Drummond et al., 2003; Duchêne et al., 2014; Ho et al., 2007, 2005; Ho and Larson, 2006). Comas and colleagues calibrated the MTBC with the human mitochondrial tree based on phylogeographic similarities and tested multiple timing models based on inferred dates of major human migrations (Comas et al., 2013). The best-performing model synchronized the emergence and spread of the MTBC with that of modern humans out of Africa approximately 70,000 years before present (BP) (Comas et al., 2013). This offered a parsimonious explanation for the global spread of tuberculosis, the seeming host-population affinities of strains within the complex, and the presence of tuberculosis (as identified by paleopathologists) in the Americas prior to European contact. However, the first ancient genomic MTBC data would challenge this view of the disease in multiple ways.

1.2.2 *Ancient DNA and tuberculosis*

As noted above, tuberculosis was the first disease to be studied in archaeogenetics via PCR (Spigelman and Lemma, 1993). Two decades worth of PCR-based aDNA studies followed, most of which targeted the IS6110 and the tuberculosis specific deletion 1 (TbD1) regions (Arriaza et al., 1995; Baker et al., 2015; Braun et al., 1998; Crubézy et al., 2006; Faerman and Jankauskas, 2000; Fletcher et al., 2003; Hershkovitz et al., 2008; Masson et al., 2013; Rothschild et al., 2001; Salo et al., 1994; Spigelman et al., 2002; Albert R. Zink et al., 2003; A. R. Zink et al., 2003; Zink et al., 2001). The findings spanned the globe and time periods from the Pleistocene to the 18th century CE (Fletcher et al., 2003). One study amplified putative MTBC DNA from bison bones excavated from Natural Trap Cave in Wyoming, USA that were dated to approximately 17,000 years before present, and the PCR results were replicated in two separate labs, which was viewed as increasing the confidence of the reported result (Rothschild et al., 2001). The PCR evidence suggests the earliest known cases of molecularly-diagnosed tuberculosis in humans are from an individual associated with the pre-domestication phase of the Dja'de el Mughara site in Syria (8800 – 8300 cal. BCE) (Baker et al., 2015) and two individuals recovered from the underwater

Neolithic site of Atlit-Yam (7300 – 6210 cal. BCE) in the eastern Mediterranean (Hershkovitz et al., 2008).

Chan and colleagues released the first NGS-based archaeogenetics study identifying MTBC from shotgun sequencing of mummified lung tissue from an 18th century Hungarian individual previously tested with PCR (2013; Fletcher, 2003). Bos and colleagues published the first ancient genomic MTBC data – enriched using hybridization capture – from pre-contact individuals from Peru, all dated to approximately 1000 years BP (Bos et al., 2014). Two remarkable findings were made from these three Peruvian MTBC genomes. First, in phylogenetic analyses, they grouped most closely with *M. pinnipedii*, a lineage of the MTBC primarily associated with seals and sea lions today. Second, Bayesian molecular dating (in which prior beliefs are combined with new data to generate updated estimates) using the ancient genomes as direct calibration points estimated the most recent common ancestor (MRCA) of the complex to have emerged no more than 6000 years BP (~4000 BCE). The following year, a follow-up publication presenting multiple MTBC genomes from the same group of Hungarian mummies analyzed by Fletcher and colleagues (2003) and Chan and colleagues (2013) was released, including a Bayesian dating analysis of Lineage 4 of the MTBC (a globally distributed human-associated lineage within the complex), which yielded similar substitution rates to the analysis presented by Bos and colleagues (2014; Kay et al., 2015). These studies, utilizing direct calibration as opposed to the indirect calibration of human migrations, reduced the estimate for the time to the MTBC MRCA by an order of magnitude. There is, of course, skepticism regarding these dates, as some challenge the use of *M. pinnipedii* genomes due to the hypothesis that the population and evolutionary dynamics of the animal-associated strains are different to those of the human-associated lineages. Manuscript A attempts to waylay this suspicion in adding a genome from a human-associated lineage to the ancient calibration points for the MTBC.

1.2.3 *Discrepancies between different lines of evidence*

Currently, the ancient genomic evidence for tuberculosis in humans does not support the hypothesis that the MTBC emerged prior to the advent of agriculture and animal domestication. As stated above, PCR evidence of MTBC DNA has been uncovered in humans (and at least one bison) prior to 6000 years before present. However, the oldest examples of PCR-amplified tuberculosis have been met with heavy scrutiny, owing in part to the genetic similarity between MTBC strains and environmental mycobacteria (Wilbur et al., 2009), which is problematic to overcome even with NGS data (Warinner et al., 2017). Additional lines of evidence, in the form of lesions identified in human bone (Canci et al., 1996; El-Najjar et al., 1996; Formicola et al., 1987; Köhler et al., 2014; Sparacello et al., 2017) and ancient lipid biomarkers (Baker et al., 2015; Lee et al., 2015, 2012), also support a more ancient origin for tuberculosis. However, diagnosing tuberculosis from lesions alone

requires vigorous differential diagnosis, as there are many osteological changes that are shared between conditions, and rigorous diagnostic practices are not always followed (Roberts and Buikstra, 2003; Wilbur et al., 2009). In the case of lipid biomarkers, similar concerns to the PCR data regarding unexplored similarities between the lipid profiles of MTBC organisms and environmental mycobacteria remain valid, and the field of ancient lipidomics itself remains undertested and underexplored (Kostyukevich et al., 2018). If all lines of evidence are ultimately shown to be valid, what could explain the discrepancy between the calibrated molecular dating and the paleopathology, PCR, and lipidomics data? In the Discussion section of Manuscript A, the possibilities are addressed in detail.

1.3 Ancient metagenomics of human microbiota

The possibility of studying microbes (modern or ancient) on the community level only became realistic after the emergence of NGS technologies. This was due directly to the subsequent reduction in cost and increase in data production, and indirectly to the expansion of genetic databases and reference sequences with which to interpret data (Mardis, 2008). Studying ancient human microbiota – the communities of microbes that inhabit the oral cavity, skin, gastrointestinal system, and vagina, making up most of the genetic diversity in the human body (Qin et al., 2010) – is essential for gaining a full understanding of human health. Associations between the oral microbiome and systemic disease have been identified (Beck and Offenbacher, 2005; Kobschull et al., 2010; Koren et al., 2011; Seymour et al., 2007). The gut microbiome is increasingly being linked to mental health and neurological impacts through, for example, the production of 90% of the human body's serotonin (Dinan et al., 2015; Schnorr et al., 2016; Yano et al., 2015). As gut microbes may be influencing human behavior through the gut-brain-axis, the gut microbiome has been shown to be influenced by human behaviors such as subsistence method (Obregon-Tito et al., 2015; Schnorr et al., 2016). These findings have emerged even as much of the microbes inhabiting human microenvironments have yet to be identified or characterized. Efforts such as the Human Microbiome Project (Consortium, 2012; Group et al., 2009) and the Human Oral Microbiome Database (Chen et al., 2010; Verma et al., 2018) have made massive strides towards characterizing and cataloguing the diversity of human microbiota. In addition, the recent development of methodologies allowing retrieval and reconstruction of genomes from uncultured (likely unculturable) microbes (Nielsen et al., 2014) has primed the field for the growth of our understanding of taxonomic diversity in human microbiota, most recently exemplified by the publication from Pasolli and colleagues (2019). As such, the volume of knowledge about the influences of human microbiota on human health and behavior is likely to grow significantly over the coming years. Feeding into and benefiting from this phenomenon is the archaeogenetics of human microbiota. So far, the ancient human

microbiome on the individual level has been mostly explored through two proxies: dental calculus for the oral microbiome and paleofeces (or coprolites) for the gut microbiome (Warinner et al., 2015).

1.3.1 *Dental calculus*

Dental calculus is a calcified biofilm that forms on the surface of human teeth. The biofilm develops through the colonization of the tooth surface by oral bacteria, which form an adhesive matrix of excreted exopolysaccharides, proteins, and extracellular DNA (Socransky and Haffajee, 2002; Warinner et al., 2014). If left undisturbed, the biofilm may calcify through the precipitation of saliva-borne calcium phosphate ions, forming a cement-like adhesion to the tooth surface (Warinner et al., 2014; White, 1997, 1991). The calcium phosphate constitution of dental calculus includes hydroxyapatite, the crystalline structure of bone, which has been shown to bind DNA exceptionally well (Purdy et al., 1996). Preus and colleagues demonstrated the presence of DNA in dental calculus through transmission electron microscopy (Preus et al., 2011). Thus, dental calculus was hypothesized to be a promising source for ancient microbial genetic data.

The first genetics studies on ancient dental calculus were released in 2013, with De La Fuente and colleagues using species-specific primers to amplify sequences from five known oral bacteria (2013), and Adler and colleagues using 16S rRNA amplicon sequencing to distinguish between the bacterial communities in ancient dental calculus spanning from the Mesolithic to the medieval period and dental tartar from modern Australians of European descent (2013). In the latter publication, the authors interpreted changes in bacterial communities over time to reflect major shifts in subsistence. The 16S rRNA gene has in the past been used as a “universal primer” for Bacteria and Archaea, but because of biases in amplification due to differences in insert lengths between regions of the 16S rRNA gene in different species, indiscriminate sequencing approaches have come into favor (Ziesemer et al., 2015). In their landmark paper, Warinner and colleagues present shotgun sequencing data from medieval European dental calculus (2014). Their presentation of more than 2,000 identified bacterial and archaeal taxa, a reconstructed genome for the oral bacterium *Tannerella forsythia*, and putative dietary DNA primed the field of archaeogenetics to look to calculus as a rich source of genetic information. However, though it had been shown that human mitochondrial genomes could be successfully enriched from dental calculus (Ozga et al., 2016), whether or not it could be used as a consistent and reliable proxy for host DNA was questionable. In addition, the preservation dynamics of DNA in dental calculus across diverse time periods, geographical locations, and individuals had not been tested. Manuscript B of this thesis (Mann et al., 2018) systematically explores diverse preservation dynamics in dental calculus as compared with paired dentin samples in a large, global dataset.

1.3.2 *Paleofeces and coprolites*

Paleofeces, the second proxy for individual ancient human gut microbes, are ancient preserved feces. In a large portion of archaeogenetics and paleoparasitology papers, the term “coprolites” is used interchangeably, though formally this refers to fossilized feces. In this thesis I will use the term paleofeces and coprolites interchangeably, though for cases in which I am referring to an existing publication, I will use the author’s chosen terminology. Paleofeces have also been used throughout the history of archaeogenetics to access host (Hofreiter et al., 2000; Poinar et al., 1998, 2001, 2003; Jenkins et al., 2012; Gilbert et al., 2008), dietary (Wood et al., 2016, 2012, 2008), and gastrointestinal parasite DNA (Iñiguez et al., 2006, 2003; Leles et al., 2008; Loreille et al., 2001). Archaeological and latrine sediments have also been used to access DNA from human-infecting gastrointestinal parasites (Côté et al., 2016; Sørensen et al., 2018). Of course, paleofeces have also been sequenced for bacterial, archaeal, and fungal gut microbiome profiles (Cano et al., 2014; Tito et al., 2012, 2008), as the ideal proxy for the ancient gut microbiome aside from preserved gastrointestinal tissue (Lugli et al., 2017; Santiago-Rodríguez et al., 2015). The micro-eukaryotic component of human microbiomes in the past, and the prevalence and impact of eukaryotic pathogens on human health, have also been relatively neglected topics (Mann, 2018). In Manuscript C, we explore composite gut microbiome and eukaryotic pathogen DNA from two medieval latrine sites in Jerusalem and Riga, Latvia.

2. Aim of the thesis

The aim of this thesis is to explore the preservation of microbial DNA in diverse archaeological materials and its utility for answering questions about human health through time. Next generation sequencing technologies have greatly expanded the possibilities of research with ancient microbial DNA, directly through the development of shotgun sequencing methods themselves and indirectly through the resultant growth of genomic reference databases. Together these allow for the considerably less biased screening of DNA extracts for relevant and informative genetic fragments. The manuscripts presented in this thesis call upon these technological and informatic developments to (a) help answer questions about the origins of the *Mycobacterium tuberculosis* complex with aDNA extracted from a calcified lung nodule, (b) systematically test the conditions of DNA preservation in dentin and dental calculus, and (c) complement and expand upon previous microscopic and immunoassay analyses of medieval latrine sediments with aDNA. In pursuing these aims, the promise and limitations of aDNA as a proxy for human health are discussed, with an emphasis on archaeological source materials for DNA. The specific questions addressed by this thesis are as follows:

Manuscript A:

- How useful are calcified nodules as a source of ancient host and pathogen DNA?
- How old is the *Mycobacterium tuberculosis* complex?
- How old is Lineage 4 of the *Mycobacterium tuberculosis* complex?
- What currently available Bayesian phylogenetic models are best suited for existing MTBC data?
- Do highly confident isolation dates for ancient genomes translate into different or more confident MRCA estimates than previously published?

Manuscript B:

- Does dental calculus act as a superior source of ancient DNA across geographical locations, time periods, human populations, and individuals?
- How do dental calculus and dentin compare as sources of gross genetic material?
- How do dental calculus and dentin compare as sources of endogenous human and microbial DNA?
- Does the preservation environment of dental calculus impact DNA differently than that of dentin?

Manuscript C:

- How do the results of microscopic and immunoassay-based identifications of eukaryotic pathogens from medieval bulk latrine sediments compare with those of metagenomic analyses?
- Can the signature of community gut microbiota be extracted from metagenomic analysis of medieval bulk latrine sediments, and how well does it preserve against environmental background?

3. Overview of manuscripts and author's contribution

3.1 Manuscript A

“A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex”

S. Sabin, A. Herbig, Å.J. Vågane, T. Ahlström, G. Bozovic, C. Arcini, D. Kühnert, and K.I. Bos

Submitted to *Genome Biology*, 29 January 2019

In Manuscript A, we extracted DNA from a calcified lung nodule that had been surgically removed from the mummified remains of Bishop Peder Winstrup of Lund. In addition to exceptional preservation of host DNA (~88% endogenous content), we found preserved DNA from the *Mycobacterium tuberculosis* complex (MTBC). After successful enrichment, we were able to reconstruct a 141-fold coverage *Mycobacterium tuberculosis* genome. We applied this genome as a calibration point to date the origin of the MTBC and Lineage 4 of the MTBC with a Bayesian approach. Using multiple models with this new calibration point, we confirmed the MTBC to be younger than 6,000 years old, supporting a Neolithic emergence for the most recent common ancestor of all extant diversity in the MTBC excluding *M. canettii*.

Note: The supplementary information file “Additional File 1,” which contains numerous large tables, can be accessed online as part of the pre-print version of this manuscript.

DOI: <https://doi.org/10.1101/588277>

Author contributions:

C. Arcini and K.I. Bos conceived of the investigation. T. Ahlström, G. Bozovic, and C. Arcini performed the exhumation and radiological analysis of the mummy and provided a paleopathological examination. G. Bozovic was responsible for the CT examinations together with imaging analysis and coordination of the calcification extraction. S. Sabin performed subsampling of the lung nodule, DNA extraction, library construction, and library amplification. S. Sabin performed preliminary metagenomic analysis and mapping of the shotgun sequencing data with aid from A. Herbig and K.I. Bos. S. Sabin performed laboratory work in preparation for in-solution capture of the UDG-treated and non-UDG-treated libraries. S. Sabin performed post-capture mapping and phylogenetic analysis with aid from Å.J. Vågane, A. Herbig, D. Kühnert, and K.I. Bos. S. Sabin, A. Herbig, D. Kühnert, and K.I. Bos. designed the molecular dating experiments. S. Sabin composed the manuscript with input

from all authors. In total, S. Sabin contributed 75% to the project, including the majority of genetic laboratory work, computational analysis, and manuscript drafting.

3.2 Manuscript B

“Differential preservation of endogenous human and microbial DNA in dental calculus and dentin”

A.E. Mann, S. Sabin, K. Ziesemer, Å.J. Vågane, H. Schroeder, A.T. Ozga, K. Sankaranarayanan, C.A. Hofman, J.A. Fellows Yates, D.C. Salazar-Garcia, B. Frohlich, M. Aldenderfer, M. Hoogland, C. Read, G.R. Milner, A.C. Stone, C.M. Lewis Jr., J. Krause, C. Hofman, K.I. Bos and C. Warinner

Published:

Mann, Sabin et al. 2018. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific Reports*, 8: 9822. DOI: 10.1038/S41598-018-28091-9

In Manuscript B, we present the first systematic exploration of ancient DNA preservation in dental calculus as compared to dentin, a commonly used source for ancient human and pathogen DNA. Dental calculus had previously been found to be an excellent genetic proxy for the human oral microbiome, but this had not been tested across geographic locations, time periods, burial contexts, and numerous individuals. We evaluated the preservation of human and microbial DNA in pairs of dentin and dental calculus across forty-eight individuals from a global sample set (n=12) spanning Iberia, Mongolia, Nepal, Guadeloupe, and Illinois, USA, and a regional sample set (n=36) from a single medieval burial ground in Ireland. We found that dental calculus preserves a higher quantity of DNA than dentin; a consistently large proportion of human-associated microbial DNA; and a small, but consistent portion of host DNA. We found that host DNA is consistently more fragmented in dental calculus than dentin. In addition, we found that some dentin samples contained about 20% oral microbiome DNA.

Note: The file containing the supplementary tables for this manuscript can be accessed online as part of the published version of this manuscript. DOI: 10.1038/S41598-018-28091-9

Author contributions:

C. Warinner and K.I. Bos designed and coordinated this study. D.C. Salazar-Garcia, B. Frohlich, M. Aldenderfer, M. Hoogland, C. Read, G.R. Milner, A.C. Stone, C.M. Lewis Jr., J. Krause and C. Hofman provided materials and resources. A. Mann, S. Sabin, K. Ziesemer, H. Schroeder, A.T. Ozga, C.A. Hofman, and D. Salazar-Garcia performed sampling, extractions, library preparation, library amplification, and extract and library quantitation for the global sample set. S. Sabin and Å.J. Vågane performed sampling, extraction, library preparation, library amplification, and extract and library quantitation for the regional sample

set. A. Mann, S. Sabin, K. Ziesemer, and J. Fellows Yates performed data analysis. A. Mann, S. Sabin, and C. Warinner composed the manuscript with input from all authors.

The original sampling, extraction, library preparation and sequencing of the “regional dataset” was performed as part of S. Sabin’s M.Sc. thesis. However, S. Sabin performed the following work specifically within the purview of this dissertation which provided new data and further analyses:

- Measurement of concentration of “regional dataset” (n=36) and “global dataset” (n=12) extracts in collaboration with A. Mann.
- Measurement of concentration of “global dataset” libraries in collaboration with A. Mann.
- Library preparation of a subset of the “regional dataset” for deep shotgun sequencing (n=13).
- Data curation of bacterial and archaeal shotgun reads from both datasets for Bray-Curtis taxonomic distance visualization, microbial source analysis, and linear discriminant analysis. S. Sabin completed the latter two analyses independently, and the Bray-Curtis taxonomic distance visualization in collaboration with A. Mann.
- Manuscript composition and figure curation in collaboration with A. Mann and C. Warinner.

In total, S. Sabin contributed 40% to the project with laboratory work, analysis, and manuscript composition.

3.3 Manuscript C

“Microbial preservation and population health inferred from metagenomic analyses of two medieval latrines”

S. Sabin, M. Ledger, H.-Y. Yeh, P. Mitchell, K.I. Bos

In preparation for submission to *Philosophical Transactions of the Royal Society*

In Manuscript C, bulk sediments from two medieval latrine sites in Riga, Latvia and Jerusalem that previously underwent microscopic and immunoassay-based analyses for the presence of eukaryotic parasites are revisited for genetic evidence of eukaryotic parasites. In contrast to other studies investigating ancient eukaryotic parasites, we (a) extracted DNA not from a concentration of filtered parasite eggs, but from the bulk sediments themselves, and (b) employed a shotgun sequencing approach instead of a targeted amplification approach. We explore possible PCR inhibition, alpha diversity measurements, possible source contributions to the total genetic content (including community gut microbiome preservation), and the preservation of eukaryotic parasite DNA. We compare the microscopy and

immunoassay findings and the metagenomics findings, discussing the value and limitations of the latter.

Author contributions:

P. Mitchell and K.I. Bos conceived of and coordinated this study. M. Ledger consulted on microscopy methodology and sample provenance. H.-Y. Yeh performed the original microscopy and immunoassay analyses on sediments from the Riga and Jerusalem latrines. S. Sabin performed sub-sampling, extraction, library preparation, and library amplification for samples from both sites. S. Sabin performed metagenomic analysis and reference-specific mapping of the resulting shotgun-sequenced data, as well as the alpha diversity and source contribution estimations. S. Sabin composed the manuscript.

In total, S. Sabin contributed 70% to the project in completing genetics laboratory work, analysis, and manuscript composition.

4. Manuscript A

A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex.

Susanna Sabin¹ (sabin@shh.mpg.de), Alexander Herbig¹ (herbig@shh.mpg.de), Åshild J. Vågane¹ (vagane@shh.mpg.de), Torbjörn Ahlström² (torbjorn.ahlstrom@ark.lu.se), Gracijela Bozovic³ (gracijela.bozovic@med.lu.se), Caroline Arcini⁴ (caroline.ahlstrom.arcini@arkeologerna.com), Denise Kühnert^{5*} (kuehnert@shh.mpg.de), Kirsten I. Bos^{1*} (bos@shh.mpg.de)

1. Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany 07745
2. Department of Archaeology and Ancient History, Lund University, Lund, Sweden 221 00
3. Department of Medical Imaging and Clinical Physiology, Skåne University Hospital Lund and Lund University, Lund, Sweden 221 00
4. Arkeologerna, National Historical Museum, Lund, Sweden 226 60
5. Transmission, Infection, Diversification & Evolution Group, Max Planck Institute for the Science of Human History, Jena, Germany 07745

*Corresponding authors

Running title: A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex

Keywords: tuberculosis, ancient DNA, *Mycobacterium tuberculosis*, molecular dating, metagenomics

ABSTRACT

Background:

Although tuberculosis accounts for the highest mortality from a bacterial infection on a global scale, questions persist regarding its origin. One hypothesis based on modern *Mycobacterium tuberculosis* complex (MTBC) genomes suggests their most recent common ancestor (MRCA) followed human migrations out of Africa ~70,000 years before present (BP). However, studies using ancient genomes as calibration points have yielded much younger MRCA dates of less than 6,000 years. Here we aim to address this discrepancy through the analysis of the highest-coverage and highest quality ancient MTBC genome available to date, reconstructed from a calcified lung nodule of Bishop Peder Winstrup of Lund (b. 1605 – d. 1697).

Results:

A metagenomic approach for taxonomic classification of whole DNA content permitted the identification of abundant DNA belonging to the human host and the MTBC, with few non-TB bacterial taxa comprising the background. Subsequent genomic enrichment enabled the reconstruction of a 141-fold coverage *M. tuberculosis* genome. In utilizing this high-quality, high-coverage 17th century *M. tuberculosis* genome as a calibration point for dating the MTBC, we employed multiple Bayesian tree models, including birth-death models, which allowed us to model pathogen population dynamics and data sampling strategies more realistically than those based on the coalescent.

Conclusions

The results of our metagenomic analysis demonstrate the unique preservation environment calcified nodules provide for DNA. Importantly, we estimate an MRCA date for the MTBC of 3258 BP (2190–4501 BP) and for Lineage 4 of 1445 BP (929–2084 BP) using multiple models, confirming a Neolithic emergence for the MTBC.

BACKGROUND

Tuberculosis, caused by organisms in the *Mycobacterium tuberculosis* complex (MTBC), has taken on renewed relevance and urgency in the 21st century due to its global distribution, its high morbidity, and the rise of antibiotic resistant strains (WHO, 2018). The difficulty in disease management and treatment, combined with the massive reservoir the pathogen maintains in human populations through latent infection (Houben and Dodd, 2016), makes tuberculosis a pressing public health challenge. Despite this, controversy exists regarding the history of the relationship between members of the MTBC and their human hosts.

Existing literature suggests two most recent common ancestor (MRCA) dates for the MTBC based on the application of Bayesian molecular dating to genome-wide *Mycobacterium tuberculosis* data. One estimate suggests the extant MTBC emerged through a bottleneck approximately 70,000 years ago, coincident with major migrations of humans out of Africa (Comas et al., 2013). This estimate was reached using exclusively modern *M. tuberculosis* genomes, with internal nodes of the MTBC calibrated by extrapolated dates for major human migrations (Comas et al., 2013). This estimate relied on congruence between the topology of MTBC and human mitochondrial phylogenies, but this congruence does not extend to human Y chromosome phylogeographic structure (Pepperell et al., 2013). As an alternative approach, the first publication of ancient MTBC genomes utilized radiocarbon dates as direct calibration points to infer mutation rates, and yielded an MRCA date for the complex of less than 6,000 years (Bos et al., 2014). This younger emergence was later supported by mutation rates estimated within the pervasive Lineage 4 (L4) of the MTBC, using four *M. tuberculosis* genomes from the late 18th and early 19th centuries (Kay et al., 2015).

Despite the agreement in studies that have relied on ancient DNA calibration so far, dating of the MTBC emergence remains controversial. Such a young age cannot account for purported detection of MTBC DNA in archaeological material that predates the MRCA estimate (e.g. Baker et al. 2015; Hershkovitz et al. 2008; Masson et al. 2013; Rothschild et

al. 2001), the authenticity of which has been challenged (Wilbur et al., 2009). Furthermore, constancy in mutation rates of the MTBC has been challenged on account of observed rate variation in modern lineages, combined with the unquantified effects of latency (Gagneux, 2018). The ancient genomes presented by Bos and colleagues, though isolated from human remains, were most closely related to *Mycobacterium pinnipedii*, a lineage of the MTBC associated with infections in seals and sea lions today (2014). Given our unfamiliarity with the demographic history of tuberculosis in sea mammal populations (Ochman et al., 1999), identical substitution rates between the pinniped lineage and human-adapted lineages of the MTBC cannot be assumed. Additionally, the identification of true genetic changes in archaeological specimens can be difficult given the similarities between MTBC and environmental mycobacterial DNA from the depositional context (Warinner et al., 2017). Though the ancient genomes published by Kay and colleagues belonged to human-adapted lineages of the MTBC, and the confounding environmental signals were significantly reduced by their funerary context in crypts, two of the four genomes used for molecular dating were derived from mixed-strain infections (2015). By necessity, diversity derived in each genome would have to be ignored for them to be computationally distinguished (Kay et al., 2015). Though ancient DNA is a valuable tool for answering the question of when the MTBC emerged, the available ancient data remains sparse and subject to case-by-case challenges.

Here, we contribute to clarifying the timing of the emergence of the MTBC and L4 using multiple Bayesian models of varying complexity through the analysis of a high-coverage 17th century *M. tuberculosis* genome extracted from a calcified lung nodule. Removed from naturally mummified remains, the nodule provided an excellent preservation environment for the pathogen, exhibiting minimal infiltration by exogenous bacteria. The nodule and surrounding lung tissue also showed exceptional preservation of host DNA, thus showing promise for this tissue type in ancient DNA investigations.

RESULTS

Pathogen identification

Computed tomography (CT) scans of the mummified remains of Bishop Peder Winstrup of Lund revealed a calcified granuloma a few millimeters (mm) in size in the collapsed right lung together with two ~5 mm calcifications in the right hilum (Figure 1). Primary tuberculosis causes parenchymal changes and ipsilateral hilar lymphadenopathy that is more common on the right side (Leung et al., 1992). Upon resolution it can leave a parenchymal scar, a small calcified granuloma (Ghon Focus), and calcified hilar nodes, which are together called a Ranke complex. In imaging this complex is suggestive of previous tuberculosis infection, although histoplasmosis can have the same appearance (Burrill et al., 2007). Histoplasmosis, however, is very rare in Scandinavia and more often seen in other parts of the world (e.g. the Americas) (Bahr et al., 2015). The imaging findings were therefore considered to result from previous primary tuberculosis. One of the calcified hilar nodes was extracted from the remains during video-assisted thoracoscopic surgery, guided by fluoroscopy. The extracted material was further subsampled for genetic analysis. DNA was extracted from the nodule and accompanying lung tissue using protocols optimized for the recovery of ancient, chemically degraded, fragmentary genetic material (Jesse Dabney et al., 2013). The metagenomic library was shotgun sequenced to a depth of approximately 3.7 million reads.

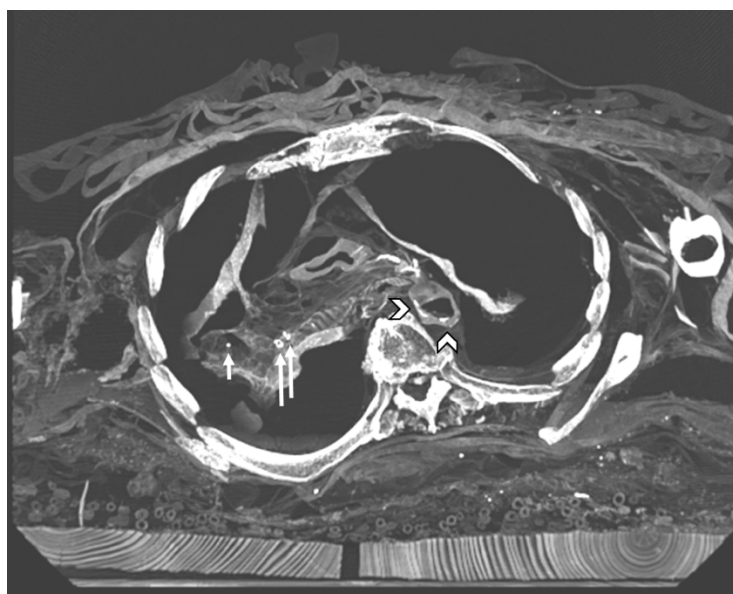


Figure 1. CT image of Ranke complex. CT image of Peder Winstrup’s chest in a slightly angled axial plane with the short arrow showing a small calcified granuloma in the probable upper lobe of the collapsed right lung, and two approximately 5 mm calcifications in the right hilum together suggesting a Ranke complex and previous primary tuberculosis. The more lateral of the two hilar calcifications was extracted for further analysis. In addition, there are calcifications in the descending aorta proposing atherosclerosis (arrowhead).

Adapter-clipped and base quality filtered reads were taxonomically binned with MALT (Vågane et al., 2018) against the full NCBI Nucleotide database (‘nt’, April 2016). In this process, 3,515,715 reads, or 95% of the metagenomic reads, could be assigned to taxa contained within the database. Visual analysis of the metagenomic profile in MEGAN6 (Huson et al., 2016) revealed the majority of these reads, 2,833,403 or 81%, were assigned to *Homo sapiens*. A further 1,724 reads assigned to the *Mycobacterium tuberculosis* complex (MTBC) node. Importantly, no other taxa in the genus *Mycobacterium* were identified, and the only other identified bacterial taxon was *Ralstonia solanacearum* (Figure 2a), a soil-dwelling plant pathogen frequently identified in metagenomic profiles of archaeological samples (Mann et al., 2018; Salanoubat et al., 2002) (Table S1 in Additional File 1).

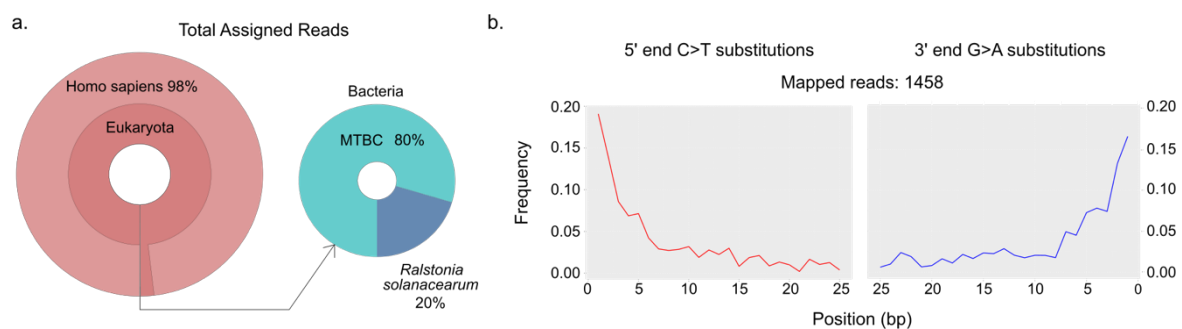


Figure 2. Screening of sequencing data from LUND1 shows preservation of host and pathogen DNA. A) Krona plots reflecting the metagenomic composition of the lung nodule. The majority of sequencing reads were aligned to *Homo sapiens* (n=2,833,403), demonstrating extensive preservation of host DNA. A small portion of reads aligned to bacterial organisms, and 80% of these reads were assigned to the MTBC node (n=1,724). B) Damage plots generated from sequencing reads mapped directly to a reconstructed MTBC

ancestor genome (Comas et al., 2010), demonstrating a pattern characteristic of ancient DNA.

Pre-processed reads were mapped to both the hg19 human reference genome and a reconstructed MTBC ancestor (TB ancestor) (Comas et al., 2010) using BWA as implemented in the Efficient Ancient Genome Reconstruction (EAGER) pipeline (Peltzer et al., 2016). Reads aligned to hg19 with direct mapping constituted an impressive 88% of the total sequencing data (Table S2 in Additional File 1). Human mitochondrial contamination was extremely low, estimated at only 1-3% using Schmutzi (Renaud et al., 2015) (Additional File 2). Reads were also mapped to the TB ancestor (Table 1). After map quality filtering and read de-duplication, 1,458 reads, or 0.045% of the total sequencing data, aligned to the reference (Table 1), and exhibited cytosine-to-thymine damage patterns indicative of authentic ancient DNA (Figure 2b) (Jesse Dabney et al., 2013; Ginolhac et al., 2011). Qualitative preservation of the tuberculosis DNA was slightly better than that of the human DNA, as the damage was greater in the latter (Table S2 in Additional File 1). Laboratory-based contamination, as monitored by negative controls during the extraction and library preparation processes, could be ruled out as the source of this DNA (Table S3 in Additional File 1).

Genomic enrichment and reconstruction

Due to the clear but low-abundance MTBC signal, a uracil DNA glycosylase (UDG) library was constructed to remove DNA lesions caused by hydrolytic deamination of cytosine residues (Briggs et al., 2010) and enriched with an in-solution capture (Fu et al., 2013; Hodges et al., 2009) designed to target genome-wide data representing the full diversity of the MTBC (see METHODS). The capture probes are based on the TB ancestor genome (Comas et al., 2010), which is equidistant from all lineages of the MTBC. The enriched library was sequenced using a paired-end, 150-cycle Illumina sequencing kit to obtain a full fragment-length distribution (Figure S1 in Additional File 2). The resulting sequencing data was then aligned to the hypothetical TB ancestor genome (Comas et al., 2010), and the

mapping statistics were compared with those from the screening data to assess enrichment (Table 1). Enrichment increased the proportion of endogenous MTBC DNA content by three orders of magnitude, from 0.045% to 45.652%, and deep sequencing yielded genome-wide data at an average coverage of approximately 141.5-fold. The mapped reads have an average fragment length of ~66 base pairs (Table 1).

Pre/post	Library treatment	Processed reads pre-mapping (n)	Unique mapped reads, quality-filtered (n)	Endogenous DNA (%)	Mean fold coverage	Mean fragment length (bp)	GC content (%)
Pre-capture	non-UDG	3696712	1458	0.045	0.018	54.31	63.89
Post-capture	UDG	59091507	9482901	45.652	141.5062	65.83	62.96

Table 1. Mapping statistics for LUND1 libraries. A comparison of the mapping statistics for the non-UDG screening library and UDG-treated MTBC enriched library of LUND1 when aligned to the MTBC ancestor genome (Comas et al., 2010). For full EAGER output, see Table S2 in Additional File 1.

We further evaluated the quality of the reconstructed genome by quantifying the amount of heterozygous positions (see METHODS). Derived alleles represented by 10-90% of the reads covering a given position with five or more reads of coverage were counted. Only 24 heterozygous sites were counted across all variant positions in LUND1. As a comparison, the other high-coverage (~125 fold) ancient genome included here – body92 from Kay et al. 2015 – contained 70 heterozygous positions.

Phylogeny and dating

Preliminary phylogenetic analysis using neighbor joining (Figures S2 and S3 in Additional File 2), maximum likelihood (Figures S4 and S5 in Additional File 2), and maximum parsimony trees (Figures S6 and S7 in Additional File 2) indicated that LUND1 groups within the L4 strain diversity of the MTBC, and more specifically, within the L4.10/PGG3 sublineage. This sublineage was recently defined by Stucki and colleagues as the clade containing L4.7, L4.8, and L4.9 (Stucki et al., 2016) according to the widely-accepted Coll nomenclature (Coll et al., 2014). Following this, we constructed two datasets to

support molecular dating of the full MTBC (Table S4 in Additional File 1) and L4 of the MTBC (Table S5 in Additional File 1), which we tested for temporal signal with TempEst (Figures S8, S9, and S10 in Additional File 2).

The dataset reflecting extant diversity of the MTBC was compiled as reported elsewhere (Bos et al., 2014), with six ancient genomes as calibration points. These included LUND1; two additional ancient genomes, body80 and body92, extracted from late 18th and early 19th century Hungarian mummies (Kay et al., 2015); and three human-isolated *Mycobacterium pinnipedii* strains from Peru (Bos et al., 2014), encompassing all available ancient *M. tuberculosis* genomes with sufficient coverage to call SNPs confidently after stringent mapping with BWA (Li and Durbin, 2009) (see METHODS; Table S4 in Additional File 1). *Mycobacterium canettii* was used as an outgroup. In generating an alignment of variant positions in this dataset, we excluded repetitive regions and regions at risk of cross-mapping with other organisms as done previously (Bos et al., 2014), as well as potentially imported sites from recombination events, which were identified using ClonalFrameML (Didelot and Wilson, 2015) (Table S6 in Additional File 1). We chose to exclude these potential recombination events despite *M. tuberculosis* being generally recognized as a largely clonal organism with no recombination or horizontal gene transfer, as these phenomena have been found to occur at low frequency in *M. canettii* (Boritsch et al., 2016). Only twenty-three variant sites were lost from the full MTBC alignment as potential imports. We called a total of 42,856 variable positions in the dataset as aligned to the TB ancestor genome. After incompletely represented sites were excluded, 11,716 were carried forward for downstream analysis.

To explore the impact of the selected tree prior and clock model, we ran multiple variations of models as available for use in BEAST2 (Bouckaert et al., 2014). We first used both a strict and a relaxed clock model together with a constant coalescent model (CC+strict, CC+UCLD). We found there to be minimal difference between the inferred rates estimated by the two models. This finding, in addition to the low rate variance estimated in all models, suggests there is little rate variation between known branches of the MTBC. Nevertheless,

the relaxed clock appeared to have a slightly better performance (Table 2). To experiment with models that allowed for dynamic populations, we applied a Bayesian skyline (SKY+UCLD) and birth-death skyline prior (BDSKY+UCLD) combined with a relaxed clock model. In the BDSKY+UCLD model, the tree was conditioned on the root. To our knowledge, this is the first instance of a birth-death tree prior being used to infer evolutionary dynamics of the MTBC while using ancient data for tip calibration. The BDSKY+UCLD model had the highest marginal likelihood value of all models applied to this dataset (Table 2).

A calibrated maximum clade credibility (MCC) tree was generated for the BDSKY+UCLD model, with 3258 years before present (BP) (95% highest posterior density [95% HPD] interval: 2190 – 4501 BP) as an estimated date of emergence for the MTBC (Figure 3a). Tree topology agrees with previously presented phylogenetic analyses of the full MTBC (Bos et al., 2014; Comas et al., 2013; O’Neill et al., 2018).

Model	Marginal Likelihood	Mean Rate (95% HPD)	Mean Rate Variance (95% HPD)	Mean Tree Height (95% HPD)
BDSKY+UCLD	- 6125044.471764 58	1.4488E-8 (9.4606E-9, 1.9632E-8)	1.6881E-17 (5.4855E-18, 3.069E-17)	3258.0478 (2189.5235, 4501.1384)
CC+UCLD	- 6126017.15694 528	1.214E-8 (7.1934E-9, 1.6448E-8)	1.2459E-17 (2.833E-18, 2.3969E-17)	4172.1961 (2585.2349, 6119.744)
SKY+UCLD	- 6127733.35000 634	1.2944E-8 (8.6149E-9, 1.7342E-8)	1.3423E-17 (4.848E-18, 2.3869E-17)	3650.4222 (2472.6434, 4992.0277)
CC+strict	- 6125541.681186 91	1.1573E-8 (8.6397E-9, 1.4509E-8)	NA	4453.1162 (3330.1516, 5619.3974)

Table 2. Model comparison for full MTBC dataset. Parameter estimates from four models applied to the full MTBC dataset: constant coalescent with uncorrelated lognormal clock (CC+UCLD), constant coalescent with strict clock (CC+strict), Bayesian skyline coalescent with uncorrelated lognormal clock (SKY+UCLD), and birth-death skyline with uncorrelated

lognormal clock (BDSKY+UCLD). Tree height is expressed in years before present, with the most recent time as 2010.

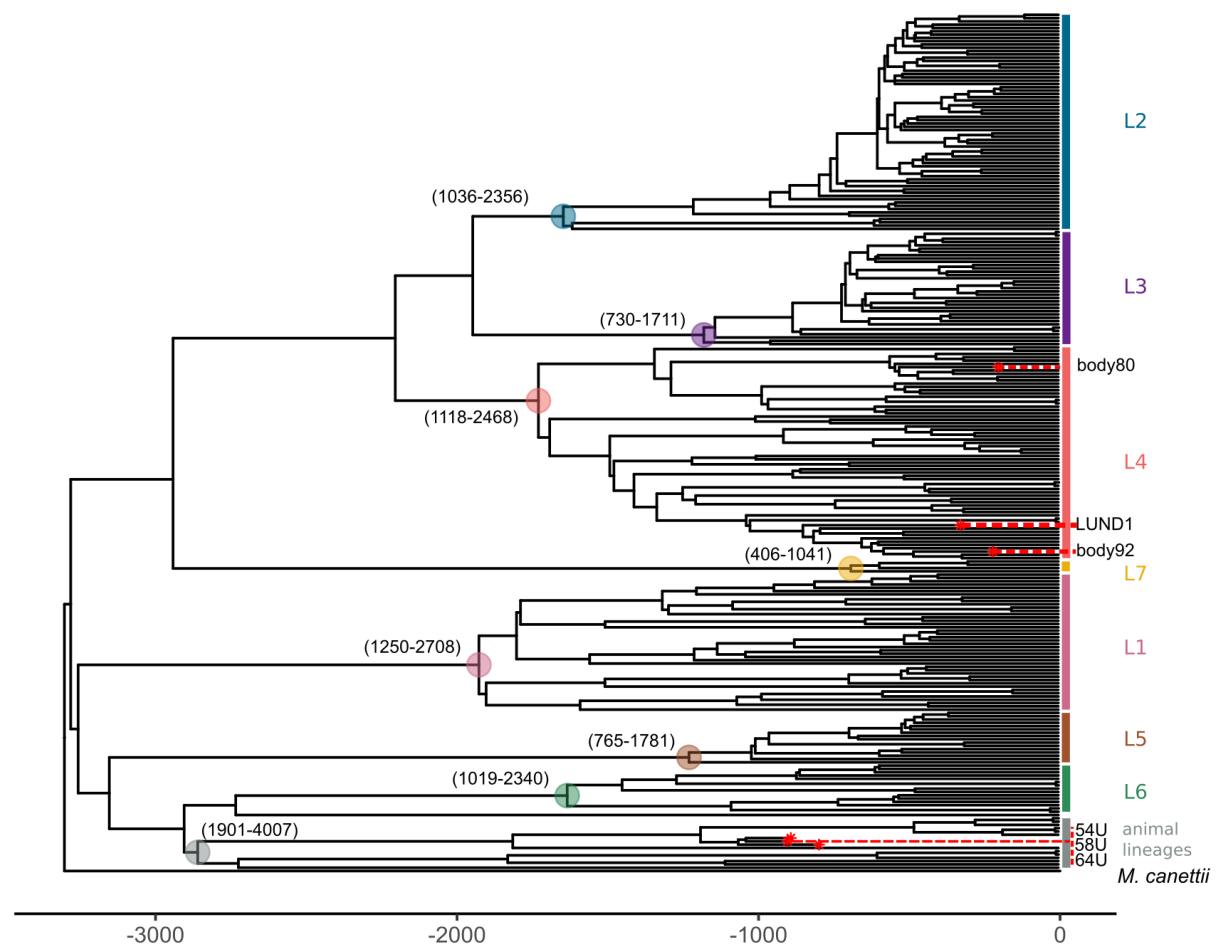


Figure 3. MTBC maximum clade credibility tree. This MCC tree of mean heights was generated from the BDSKY+UCLD model as applied to the full MTBC dataset. Lineages are labeled on the right side. The ancient genomes are indicated by red asterisks and labeled on the side with their sample names. The outgroup is labeled as “*M. canettii*.” The 95% HPD intervals of the heights of nodes ancestral to each lineage are indicated as (lower boundary - upper boundary) in years before present. Ancestral nodes are highlighted by a circle colored to match the lineage label. The time scale is expressed as years before present, with the most recent time as 2010.

The L4 dataset includes LUND1 and the two Hungarian mummies described above (Kay et al., 2015) as calibration points. We selected 149 modern genomes representative of the known diversity of L4 from previously published datasets (Additional File 2) (Comas et

al., 2013, 2010; Stucki et al., 2016). A modern Lineage 2 (L2) genome was used as an outgroup. After the exclusion of sites as discussed above (Table S7 in Additional File 1), a SNP alignment of these genomes in reference to the reconstructed TB ancestor genome (Comas et al., 2010) included a total of 17,333 variant positions, excluding positions unique to the L2 outgroup. Only fifteen variant sites were lost from the L4 dataset alignment. After sites missing from any alignment in the dataset were excluded from downstream analysis, 10,009 SNPs remained for phylogenetic inference. A total of 810 SNPs were identified in LUND1, of which 126 were unique to this genome. A SNP effect analysis (Cingolani et al., 2012) was subsequently performed on these derived positions (Additional File 2; Table S8 in Additional File 1).

We applied the same models as described above for the full MTBC dataset, with the addition of a birth-death skyline model conditioned on the origin of the root (BDSKY+UCLD+origin). All mean tree heights are within 250 years of each other and the 95% HPD intervals largely overlap. The BDSKY+UCLD and BDSKY+UCLD+origin models show the highest marginal likelihood values after stepping stone sampling. We employed the BDSKY+UCLD+origin model to determine if the estimated origin of the L4 dataset agreed with the tree height estimates for the full MTBC dataset. Intriguingly, the estimated origin parameter (Table 3), or the ancestor of the tree root, largely overlaps with the 95% HPD range for MTBC tree height as seen in Table 2.

A calibrated MCC tree (Figure 4) was generated based on the BDSKY+UCLD model for the L4 dataset. This model yielded an estimated date of emergence for L4 of 1445 BP (95% HPD: 929-2084 BP). The tree reflects the ten-sublineage topology presented by Stucki and colleagues (2016), with LUND1 grouping with the L4.10/PGG3 sublineage.

Model	Marginal Likelihood	Mean Rate (95% HPD)	Mean Rate Variance (95% HPD)	Mean Tree Height (95% HPD)	Origin (BDSKY only)
BDSKY+UCLD	-6033864.2003	3.1885E-8 (1.9488E-8, 4.4007E-8)	4.991E-17 (1.0674E-17, 8.9835E-17)	1444.5416 (929.3966, 2083.7636)	NA
BDSKY+UCLD+origin	-60327945.1483	3.4761E-8 (2.447E-8, 4.5029E-8)	5.5123E-17 (1.9718E-17, 9.4555E-17)	1319.2463 (952.8702, 1761.4382)	2310.916(1165.2155, 3372.9253)
CC+UCLD	-6043356.1504	3.1068E-8 (1.988E-8, 4.1624E-8)	4.3865E-17 (1.3291E-17, 7.806E-17)	1569.0512 (1054.607, 2225.4758)	NA
SKY+UCLD	-6034698.362	2.8097E-8 (1.5329E-8, 3.9927E-8)	3.7609E-17 (6.0593E-18, 7.1919E-17)	1690.536 (1016.2712, 2646.5163)	NA
CC+strict	-6034091.5119	2.9299E-8 (2.2173E-8, 3.6637E-8)	NA	1567.544 (1186.1186, 1978.6488)	NA

Table 3. Model comparison for L4 dataset. Selected parameter estimates from five models

applied to the Lineage 4 dataset: constant coalescent with uncorrelated lognormal clock (CC+UCLD), constant coalescent with strict clock (CC+strict), Bayesian skyline coalescent with uncorrelated lognormal clock (SKY+UCLD), birth-death skyline with uncorrelated lognormal clock and tree conditioned on the root (BDSKY+UCLD), and birth-death skyline with uncorrelated lognormal clock with origin parameter estimate (BDSKY+UCLD+origin). Tree height is expressed in years before present, with the most recent time as 2010.

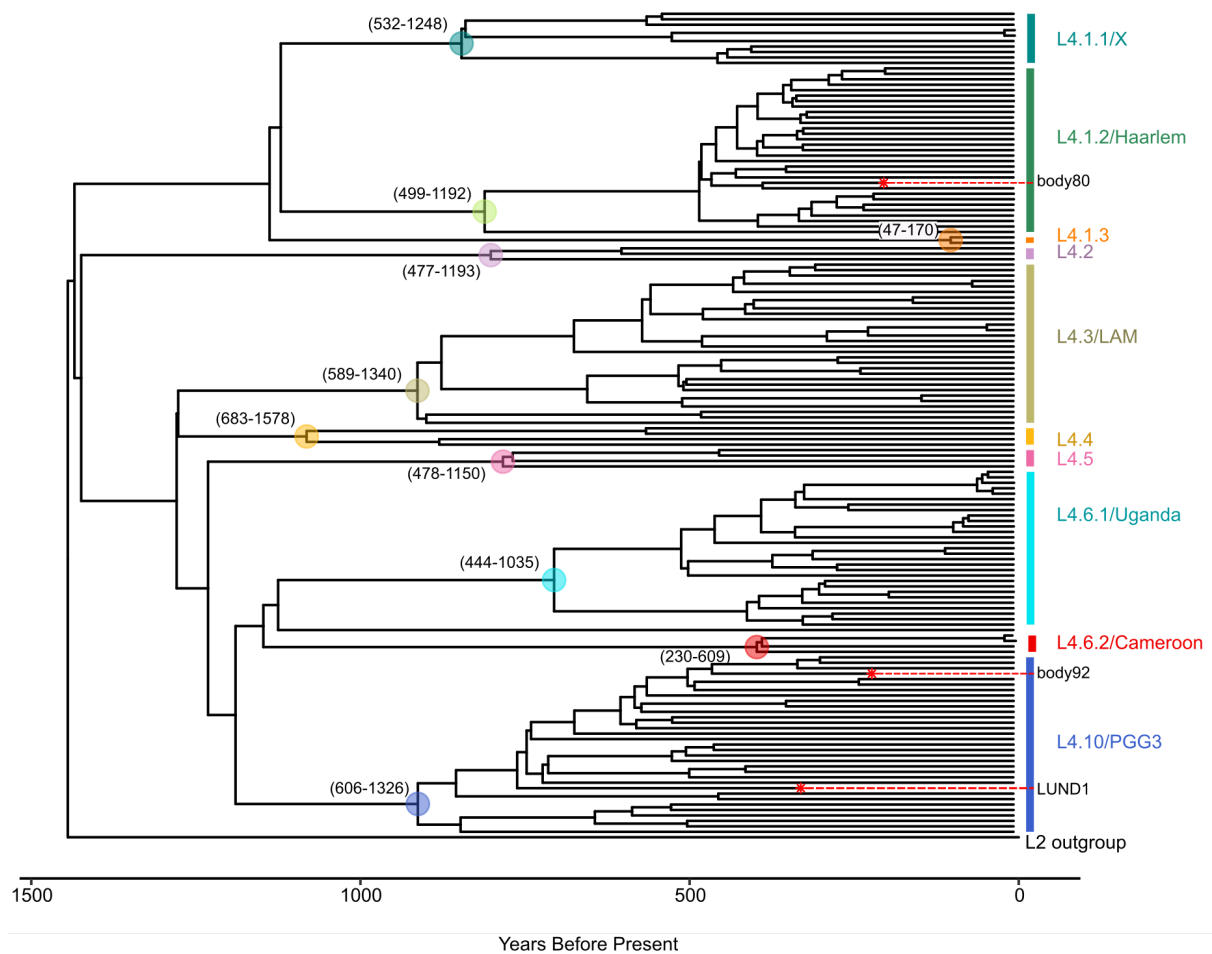


Figure 4. L4 maximum clade credibility tree. This MCC tree of mean heights was generated from the BDSKY+UCLD model as applied to the L4 dataset. Sublineages are labelled on the right side. The ancient genomes are indicated by red asterisks and labelled with their sample name. The Lineage 2 outgroup, represented by L2_N0020, is labeled on the side. The 95% HPD interval for node height is displayed for ancestral nodes of each sublineage as (lower boundary - upper boundary) in years before present. Ancestral nodes are highlighted by a circle colored to match the sublineage label. The time scale is expressed as years before present, with the most recent time as 2010.

DISCUSSION

The increasing number of ancient *Mycobacterium tuberculosis* genomes is steadily reducing the uncertainty of molecular dating estimates for the emergence of the MTBC. Here, using the ancient data available to date, we directly calibrate the MTBC time tree, and

confirm that known diversity within the complex is derived from a common ancestor that existed ~2000-6000 years before present (Figure 3; Table 2) (Bos et al., 2014; Kay et al., 2015). Our results support the hypothesis that the MTBC emerged during the Neolithic, and not before. The Neolithic revolution generally refers to the worldwide transition in lifestyle and subsistence from more mobile, foraging economies to more sedentary, agricultural economies made possible by the domestication of plants and animals. The period during which it occurred varies between regions. In Africa, where the MTBC is thought to have originated (Comas et al., 2013; Gutierrez et al., 2005; Hershberg et al., 2008; Wirth et al., 2008), the spread of animal domestication in the form of pastoralism appears to have its focus around ~3000 BCE, or 5000 BP, across multiple regions (Shillington, 2012). The estimates presented here place the emergence of tuberculosis amidst the suite of human health impacts that took place as a consequence of the Neolithic lifestyle changes often referred to collectively as the first epidemiological transition (Armstrong et al., 2005; Cohen and Armstrong, 1984).

Tuberculosis has left testaments to its history as a human pathogen in the archaeological record (Roberts and Buikstra, 2003), and some skeletal evidence has implied the existence of tuberculosis in humans and animals pre-dating the upper 95% HPD boundary for the MTBC MRCA presented here (Baker et al., 2015; Canci et al., 1996; El-Najjar et al., 1996; Formicola et al., 1987; Hershkovitz et al., 2008; Köhler et al., 2014; Rothschild et al., 2001; Sparacello et al., 2017). However, it is important to explore the evolutionary history of the MTBC through molecular data. Furthermore, it is crucial to base molecular dating estimates on datasets that include ancient genomes, which expand the temporal sampling window and provide data from the pre-antibiotic era. Numerous studies have found long-term nucleotide substitution rate estimates in eukaryotes and viruses to be dependent on the temporal breadth of the sampling window, and it is reasonable to assume the same principle applies to bacteria (Achtman, 2016; Drummond et al., 2003; Duchêne et al., 2014; Ho et al., 2007, 2005; Ho and Larson, 2006). Additionally, rate variation over time and between lineages, which may arise due to changing evolutionary dynamics such as

climate and host biology, can impact the constancy of the molecular clock (Achtman, 2016; Duchêne et al., 2014). Though models have been developed to accommodate uncertainty regarding these dynamics (Drummond et al., 2006), temporally structured populations can provide evidence and context for these phenomena over time and can aid researchers in refining models appropriate for the taxon in question (Drummond et al., 2003).

In addition to our MRCA estimate for the MTBC, we present one for L4, which is among the most globally dominant lineages in the complex (Brynildsrud et al., 2018; Stucki et al., 2016). Our analyses yielded MRCA dates between ~900-2500 years before present, as extrapolated from the 95% HPD intervals of all models (Table 3), with the mean dates spanning from 320-691 CE. These results are strikingly similar to those found in two prior publications, and support the idea proposed by Kay and colleagues that L4 may have emerged during the late Roman period (Bos et al., 2014; 2015). However, there exist discrepancies between different estimates for the age of this lineage in available literature that overlap with the upper (O'Neill et al., 2019) and lower (Brynildsrud et al., 2018) edges of the 95% HPD intervals reported here. In addition, recent phylogeographic analyses of the MTBC and its lineages had ambiguous results for L4, with the internal nodes being assigned to either African or European origins depending on the study or different dataset structures used within the same study (Brynildsrud et al., 2018; O'Neill et al., 2019). This finding belies a close relationship between ancestral L4 strains in Europe and Africa (Brynildsrud et al., 2018; O'Neill et al., 2019). Stucki and colleagues delineated L4 into two groups based on the extent of their geographic distribution: globally distributed “generalist” sublineages and highly local “specialist” sublineages that do not appear outside a restricted geographical niche (Stucki et al., 2016). Thus far, the “specialist” sublineages are found regionally on the African continent. A clear phylogenetic relationship explaining the distinction between geographically expansive and limited strains has not been established. Specifically, LUND1 falls within the globally distributed, “generalist” L4.10/PGG3 sublineage that shares a clade with two “specialist” sublineages: L4.6.1/Uganda and L4.6.2/Cameroon (Figure 4) (Stucki et al., 2016). Elucidating the phenomenon that separated L4.10/PGG3 and the L4.6 lineages could

offer relevant clues about the evolutionary relationship between specific populations of MTBC organisms and specific populations of humans by selection or genetic drift discussed previously (Gagneux et al., 2006; Hershberg et al., 2008). Assuming modern L4 diversity in Africa was driven by colonial migrations from Europe to Africa (Brynildsrud et al., 2018), why do we not see the L4.6 lineages more frequently in European populations as we do their sister clade? However, the current discrepancies over the age and geographic origin of L4 make interpretations of existing data unreliable for questions of such specificity and complexity at this time. These discrepancies could be due to differences in genome selection, SNP selection, and/or model selection and parameterization. Until more diverse, high-quality ancient L4 genomes are generated, creating a more temporally and geographically structured dataset, it is unlikely we will gain clarity.

Going deeper into comparisons between the results presented here and those from prior studies, mutation rate estimates in the L4 and full MTBC analyses were lower than previous estimates for comparable datasets, but within the same order of magnitude, with all mean and median estimates ranging between $1\text{E-}8$ and $5\text{E-}8$ (Bos et al., 2014; Kay et al., 2015) (Table 2). Nucleotide substitution rates inferred based on modern tuberculosis data are close to, but slightly higher than those based on ancient calibration, with multiple studies finding rates of approximately $1\text{E-}7$ substitutions per site per year in multiple studies (Ford et al., 2011; Pepperell et al., 2013). Despite a strict clock model having been rejected by the MEGA-CC molecular clock test (Kumar et al., 2012) for both the L4 and full MTBC datasets, the clock rate variation estimates do not surpass $9\text{E-}17$ in any model. Additionally, there is little difference between the clock rates estimated in the L4 and full MTBC datasets suggesting the rate of evolution in L4 does not meaningfully differ from that of the full complex (Tables 2 and 3; Figure 5; Figures S12 and S13 in Additional File 2).

Importantly, we explored our data through multiple models, including birth-death tree priors. In our opinion, these models offer more robust parameterization options for heterochronous datasets that are unevenly distributed over time, such as those presented here, by allowing for uneven sampling proportions across different time intervals of the tree

(Stadler et al., 2013). Recent studies have demonstrated the importance of selecting appropriate tree priors for the population under investigation, as well as the differences between birth-death and coalescent tree priors (Boskova et al., 2018; Möller et al., 2018). It is notable that the estimates reported here roughly agree across multiple demographic and clock models implemented in BEAST2. The estimate of the origin height for the L4 dataset as calculated with the birth-death Skyline model overlaps with the 95% HPD intervals for the tree height estimates across models in the full MTBC dataset.

In addition to confirming the findings of prior publications, this study contributes a high-coverage, contamination-free, and securely dated ancient *M. tuberculosis* genome for future dating efforts, which may include more ancient data or more realistic models. Much of this quality likely comes from the unique preservation environment of the calcified nodule. In the case of tuberculosis, such nodules form from host immunological responses in the waning period of an active pulmonary infection (Brown et al., 1994) and remain in lung tissue, characterizing the latent form of the disease. Host immune cells were likely responsible for the dominant signal of 89% human DNA in the LUND1 metagenomic screening library. Similar levels of preservation have been observed through analyses of ancient nodules yielding *Brucella* (Kay et al, 2014) and urogenital bacterial infections (DeVault et al, 2017), with pathogen preservation surpassing what we report here.

LUND1 avoided multiple quality-related problems often encountered in the identification and reconstruction of ancient genetic data from the MTBC. The genome is of high quality both in terms of its high coverage and low heterozygosity. Despite the low quantity of MTBC DNA detected in the preliminary screening data, in-solution capture enriched the proportion of endogenous DNA by three orders of magnitude (Table 1). The resultant genomic coverage left few ambiguous positions at which multiple alleles were represented by greater than 10% of the aligned reads. This extremely low level of heterozygosity indicated that LUND1 contained a dominant signal of only one MTBC strain. This circumvented analytical complications that can arise from the simultaneous presence of multiple MTBC strains associated with mixed infections, or from the presence of abundant

non-MTBC mycobacteria stemming from the environment. The preservation conditions of Bishop Winstrup's remains, mummified in a crypt far from soil, left the small MTBC signal unobscured by environmental mycobacteria or by the dominance of any other bacterial organisms (Figure 2a). The unprecedented quality of LUND1 and the precision of its calibration point (historically recorded year of death) made it ideal for Bayesian molecular dating applications.

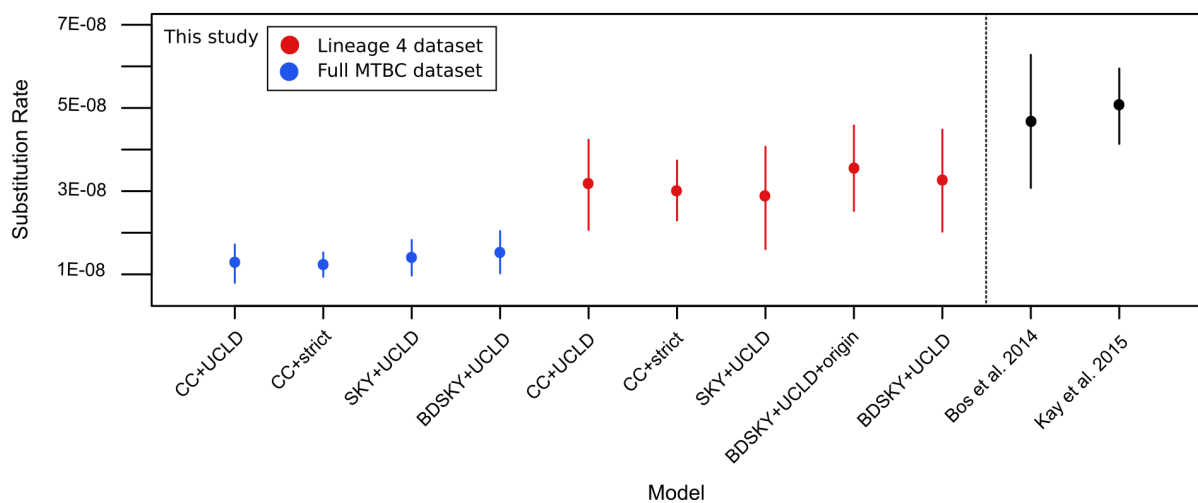


Figure 5. Substitution rate comparison across models and studies. Mean substitution rate per site per year for all models is expressed by a filled circle, with extended lines indicating the 95% HPD interval for that parameter. The Bos et al. 2014 and Kay et al. 2015 ranges are based on the reported rate values in each study. The Bos et al. 2014 range is based on a full MTBC dataset, while the Kay et al. 2015 range is based on an L4 dataset. All values presented here fall within one order of magnitude.

As the practice of applying ancient genomic data is still in its nascent stages, there are caveats to the results of this study. First, this analysis excludes diversity within *M. canettii* – a bacterium that can cause pulmonary tuberculosis – from the MTBC dataset, and as such our estimate does not preclude the possibility of a closely related ancestor having caused infections indistinguishable from tuberculosis in humans before 6,000 BP. The inferred MRCA could be restricted to a lineage that survived an evolutionary bottleneck or selective

sweep, possibly connected to its virulence in humans as suggested elsewhere, albeit as a considerably more ancient event (Jankute et al., 2017; Sreevatsan et al., 1997). It is possible there were pathogenic sister lineages to the MTBC that existed prior to this reduction in diversity and are not represented by extant MTBC diversity. Additionally, despite the use of ancient data, our temporal sampling window is still narrow given the estimated age of the MTBC and L4. For the MTBC dataset no samples pre-date 1,000 years before present, and for L4, no samples predate 350 years before present. It could be argued the ancient L4 genomes available to date represent samples taken in the midst of an epidemic – namely, the “White Plague” of tuberculosis, which afflicted Europe between the 17th and 19th centuries (Dubos and Dubos, 1952). For a slow-evolving bacterial pathogen like tuberculosis, it is possible our sampling window of ancient genomes is subject to the very issue they are meant to alleviate: the time-dependency of molecular clocks (Achtman, 2016; Duchêne et al., 2014; Ho et al., 2007, 2005). The genomes sampled from pre-contact Peruvian remains do not derive from a known epidemic period in history and add temporal spread to our MTBC dataset, but also belong to a clade of animal-associated strains (*M. pinnipedii*) that may have been subject to dramatically different evolutionary pressures compared to the human-associated lineages of the complex due to differing host biology and population dynamics. However, our use of a relaxed clock model allowed for the estimation and accommodation of variable rates across different branches of the complex. We do not see evidence for divergent substitution rates among the branches leading to the Peruvian *M. pinnipedii* strains (Figure S12 in Additional File 2). On a related matter, we may be missing diversity for some lineages (e.g. L6, L7, animal lineages) for which whole genome data is sparse. The available ancient MTBC genomes also suffer from a lack of lineage diversity, with only pinniped strains and L4 represented.

Filling the MTBC time tree with more ancient genomes from diverse time periods, locations, and lineages would address the limitations listed above. The most informative data would a) derive from an Old World context (i.e. Europe, Asia, or Africa) pre-dating the White Plague in Europe or b) come from any geographical location or pre-modern time period, but

belong to one of the MTBC lineages not yet represented by ancient data. An ideal data point, which would clarify many open questions and seeming contradictions related to the evolutionary history of the MTBC, would derive from Africa, the inferred home of the MTBC ancestor (Comas et al., 2013; Gutierrez et al., 2005; Hershberg et al., 2008; Wirth et al., 2008), and pre-date 2,000 years before present. A genome of this age would test the lower boundaries of the 95% HPD tree height intervals estimated in the full MTBC models presented here. Until recently it would have been considered unrealistic to expect such data to be generated from that time period and location. Innovations and improvements in ancient DNA retrieval and enrichment methods, however, have brought this expectation firmly into the realm of the possible (Fu et al., 2013; Gansauge et al., 2017). Ancient bacterial pathogen genomes have now been retrieved from remains from up to 5,000 years before present (Andrades Valtueña et al., 2017; Rasmussen et al., 2015; Spyrou et al., 2018b) and recent studies have reported the recovery of human genomes from up to 15,000 year-old remains from north Africa (Loosdrecht et al., 2018; Schuenemann et al., 2017).

CONCLUSIONS

Here we offer confirmation that the extant MTBC, and all available ancient MTBC genomes, stem from a common ancestor that existed a maximum of 6,000 years before present. Many open questions remain, however, regarding the evolutionary history of the MTBC and its constituent lineages, as well as the role of tuberculosis in human history. Elucidating these questions is an iterative process, and progress will include the generation of diverse ancient *M. tuberculosis* genomes, and the refinement and improved parameterization of Bayesian models that reflect the realities of MTBC (and other organisms') population dynamics and sampling frequencies over time. To aid in future attempts to answer these questions, this study provides an ancient MTBC genome of impeccable quality and explores the first steps in applying birth-death population models to modern and ancient TB data.

METHODS

Lung nodule identification

The paleopathological investigation of the body of Winstrup is based on extensive CT-scan examinations with imaging of the mummy and its bedding performed with a Siemens Somatom Definition Flash, 128 slice at the Imaging Department of Lund University Hospital. Ocular inspection of the body other than of the head and hands was not feasible, since Winstrup was buried in his episcopal robes and underneath the body was wrapped in linen strips. The velvet cap and the leather gloves were removed during the investigation. The body was naturally mummified and appeared to be well preserved with several internal organs identified.

The imaging was quite revealing. The intracranial content was lost with remains of the brain in the posterior skull base. Further, the dental status was poor with several teeth in the upper jaw affected by severe attrition, caries and signs of tooth decay, as well as the absence of all teeth in the lower jaw. Most of the shed teeth were represented by closed alveoli, indicating antemortem tooth loss. Along with the investigation of the bedding, a small sack made of fabric was found behind the right elbow containing five teeth: two incisors, two premolars and one molar. The teeth in the bag complemented the remaining teeth in the upper jaw. It is feasible that the teeth belonged to Winstrup and were shed several years before he died. A fetus approximately five months of age was also found in the bedding, underneath his feet.

Both lungs were preserved but collapsed with findings of a small parenchymal calcification and two ~5 mm calcifications in the right hilum (Figure 1). The assessment was that these could constitute a Ranke complex, suggestive of previous primary tuberculosis (Brown et al., 1994). A laparoscopy was performed at the Lund University Hospital in a clinical environment whereby the nodules were retrieved. Furthermore, several calcifications were also found in the aorta and the coronary arteries, suggesting the presence of atherosclerosis. The stomach, liver and gall bladder were preserved, and several small gallstones were observed. The spleen could be identified but not the kidneys. The intestines

were there, however, collapsed except for the rectum that contained several large pieces of concretions. The bladder and the prostate could not be recognized.

The skeleton showed several pathological changes. Findings on the vertebrae consistent with DISH (Diffuse idiopathic skeletal hyperostosis) were present in the thoracic and the lumbar spine. Reduction of the joint space in both hip joints and the left knee joint indicate that Winstrup was affected by osteoarthritis. No signs of gout or osteological tuberculosis (i.e. Pott's disease) were found.

Neither written sources nor the modern examination of the body of Winstrup reveal the immediate cause of death. However, it is known that he was bedridden for at least two years preceding his death. Historical records indicate that gallstones caused him problems while travelling to his different parishes. Additionally, he was known to have suffered from tuberculosis as a child, which may have recurred in his old age.

Sampling and extraction

Sampling of the lung nodule, extraction, and library preparation were conducted in dedicated ancient DNA clean rooms at the Max Planck Institute for the Science of Human History in Jena, Germany. The nodule was broken using a hammer, and a 5.5 mg portion of the nodule was taken with lung tissue for extraction according to a previously described protocol with modifications (Jesse Dabney et al., 2013). The sample was first decalcified overnight at room temperature in 1 mL of 0.5 M EDTA. The sample was then spun down, and the EDTA supernatant was removed and frozen. The partially decalcified nodule was then immersed in 1 mL of a digestion buffer with final concentrations of 0.45 M EDTA and 0.25 mg/mL Proteinase K (Qiagen) and rotated at 37°C overnight. After incubation, the sample was centrifuged. The supernatants from the digestion and initial decalcification step were purified using a 5 M guanidine-hydrochloride binding buffer with a High Pure Viral Nucleic Acid Large Volume kit (Roche). The extract was eluted in 100 µl of a 10mM tris-hydrochloride, 1 mM EDTA (pH 8.0), and 0.05% Tween-20 buffer (TET). Two negative controls and one positive control sample of cave bear bone powder were processed

alongside LUND1 to control for reagent/laboratory contamination and process efficiency, respectively.

Library preparation and shotgun screening sequencing

Double-stranded Illumina libraries were constructed according to an established protocol with some modifications (Meyer and Kircher, 2010). Overhangs of DNA fragments were blunt-end repaired in a 50 μ l reaction including 10 μ l of the LUND1 extract, 21.6 μ l of H₂O, 5 μ l of NEB Buffer 2 (New England Biolabs), 2 μ l dNTP mix (2.5 mM), 4 μ l BSA (10 mg/ml), 5 μ l ATP (10 mM), 2 μ l T4 polynucleotide kinase, and 0.4 μ l T4 polymerase, then purified and eluted in 18 μ l TET. Illumina adapters were ligated to the blunt-end fragments in a reaction with 20 μ l Quick Ligase Buffer, 1 μ l of adapter mix (0.25 μ M), and 1 μ l of Quick Ligase. Purification of the blunt-end repair and adapter ligation steps was performed using MinElute columns (Qiagen). Adapter fill-in was performed in a 40 μ l reaction including 20 μ l adapter ligation eluate, 12 μ l H₂O, 4 μ l Thermopol buffer, 2 μ l dNTP mix (2.5 mM), and 2 μ l Bst polymerase. After the reaction was incubated at 37°C for 20 minutes, the enzyme was heat deactivated with a 20 minute incubation at 80°C. Four library blanks were processed alongside LUND1 to control for reagent/laboratory contamination. The library was quantified using a real-time qPCR assay (Lightcycler 480 Roche) with the universal Illumina adapter sequences IS7 and IS8 as targets. Following this step, the library was double indexed (Kircher et al., 2012) with a unique pair of indices over two 100 μ l reactions using 19 μ l of template, 63.5 μ l of H₂O, 10 μ l PfuTurbo buffer, 1 μ l PfuTurbo (Agilent), 1 μ l dNTP mix (25mM), 1.5 μ l BSA (10 mg/ml), and 2 μ l of each indexing primer (10 μ M). The master mix was prepared in a pre-PCR clean room and transported to a separate lab for amplification. The two reactions were purified and eluted in 25 μ l of TET each over MinElute columns (Qiagen), then assessed for efficiency using a real-time qPCR assay targeting the IS5 and IS6 sequences in the indexing primers. The reactions were then pooled into one double-indexed library. Approximately one-third of the library was amplified over three 70 μ l PCR reactions using 5 μ l of template each and Herculase II Fusion DNA Polymerase (Agilent). The products were MinElute purified, pooled, and quantified using an Agilent Tape Station

D1000 Screen Tape kit. LUND1 and the corresponding negative controls were sequenced separately on an Illumina NextSeq 500 using single-end, 75-cycle, high-output kits.

Pathogen identification and authentication

De-multiplexed sequencing reads belonging to LUND1 were processed *in silico* with the EAGER pipeline (v.1.92) (Peltzer et al., 2016). ClipAndMerge was used for adapter removal, fragment length filtering (minimum sequence length: 30 bp), and base sequence quality filtering (minimum base quality: 20). MALT v. 038 (Vågane et al., 2018) was used to screen the metagenomic data for pathogens using the full NCBI Nucleotide database ('nt', April 2016) with a minimum percent identity of 85%, a minSupport threshold of 0.01, and a topPercent value of 1.0. The resulting metagenomic profile was visually assessed with MEGAN6 CE (Huson et al., 2016). The adapter-clipped reads were additionally aligned to a reconstructed MTBC ancestor genome (Comas et al., 2010) with BWA (Li and Durbin, 2009) as implemented in EAGER (-l 1000, -n 0.01, -q 30). Damage was characterized with DamageProfiler in EAGER (Neukamm and Peltzer, 2018).

In-solution capture probe design

Single-stranded probes for in-solution capture were designed using a computationally extrapolated ancestral genome of the MTBC (Comas et al., 2010). The probes are 52 nucleotides in length with a tiling density of 5 nucleotides, yielding a set of 852,164 unique probes after the removal of duplicate and low complexity probes. The number of probes was raised to 980,000 by a random sampling among the generated probe sequences. A linker sequence (5'-CACTGCGG-3') was attached to each probe sequence, resulting in probes of 60 nucleotides in length, which were printed on a custom-design 1 million-feature array (Agilent). The printed probes were cleaved off the array, biotinylated and prepared for capture according to Fu et al. (2013).

UDG library preparation and in-solution capture

Fifty microliters of the original LUND1 extract were used to create a uracil-DNA glycosylase (UDG) treated library, in which the post-mortem cytosine to uracil modifications, which cause characteristic damage patterns in ancient DNA, are removed. The template

DNA was treated in a buffer including 7 μ l H₂O, 10 μ l NEB Buffer 2 (New England Biolabs), 12 μ l dNTP mix (2.5 mM), 1 μ l BSA (10 mg/ml), 10 μ l ATP (10 mM), 4 μ l T4 polynucleotide kinase, and 6 μ l USER enzyme (New England Biolabs). The reaction was incubated at 37°C for three hours, then 4 μ l of T4 polymerase was added to the library to complete the blunt-end repair step. The remainder of the library preparation protocol, including double indexing, was performed as described above.

The LUND1 UDG-treated library was amplified over two rounds of amplification using Herculase II Fusion DNA Polymerase (Agilent). In the first round, five reactions using 3 μ l of template each were MinElute purified and pooled together. The second round of amplification consisted of three reactions using 3 μ l of template each from the first amplification pool. The resulting products were MinElute purified and pooled together. The final concentration of 279 ng/ μ l was measured using an Agilent Tape Station D1000 Screen Tape kit (Agilent). A portion of the non-UDG library (see above) was re-amplified to 215 ng/ μ l. A 1:10 pool of the non-UDG and UDG amplification products was made to undergo capture. A pool of all associated negative control libraries (Supplementary Table 2) and a positive control known to contain *M. tuberculosis* DNA also underwent capture in parallel with the LUND1 libraries. Capture was performed according to an established protocol (Hodges et al., 2009), and the sample product was sequenced on an Illumina HiSeq 4000 with a 150-cycle paired end kit to a depth of ~60 million paired reads. The blanks were sequenced on a NextSeq 500 with a 75-cycle paired end kit.

Genomic reconstruction, heterozygosity, and SNP calling

For the enriched, UDG-treated LUND1 sequencing data, de-multiplexed paired-end reads were processed with the EAGER pipeline (v. 1.92) (Peltzer et al., 2016), adapter-clipped with AdapterRemoval, and aligned to the MTBC reconstructed ancestor genome with in-pipeline BWA (-l 32, -n 0.1, -q 37). Previously published ancient and modern *Mycobacterium tuberculosis* genomic data (Supplementary Table 4, Supplementary Table 5) were processed as single-end sequencing reads, but otherwise processed identically in the EAGER pipeline. Genome Analysis Toolkit (GATK) UnifiedGenotyper was used to call SNPs

using default parameters and the EMIT ALL SITES output option (Auwera et al., 2013). We used MultiVCFAnalyzer (v0.87 <https://github.com/alexherbig/MultiVCFAnalyzer>) (Bos et al., 2014) to create and curate SNP alignments for the L4 (Supplementary Table 5) and full MTBC (Supplementary Table 4) datasets based on SNPs called in reference to the TB ancestor genome (Comas et al., 2010), with repetitive sequences, regions subject to cross-species mapping, and potentially imported sites excluded. The repetitive and possibly cross-mapped regions were excluded as described previously (Bos et al., 2014). Potentially imported sites were identified using ClonalFrameML (Didelot and Wilson, 2015) separately for each dataset, using full genomic alignments and trees generated in RAxML (Stamatakis, 2014) as input without the respective outgroups. Remaining variants were called as homozygous if they were covered by at least 5 reads, had a minimum genotyping quality of 30, and constituted at least 90% of the alleles present at the site. Outgroups for each dataset were included in the SNP alignments, but no variants unique to the selected outgroup genomes were included. Minority alleles constituting over 10% were called and assessed for LUND1 to check for a multiple strain *M. tuberculosis* infection. Sites with missing or incomplete data were excluded from further analysis.

Phylogenetic analysis

Neighbor joining (Figures S2 and S3 in Additional File 2), maximum likelihood (Figures S4 and S5 in Additional File 2), and maximum parsimony (Figures S6 and S7 in Additional File 2) trees were generated for the L4 and full MTBC datasets (Tables S4 and S5 in Additional File 1), with 500 bootstrap replications per tree. Maximum parsimony and neighbor joining trees were configured using MEGA-Proto and executed using MEGA-CC (Kumar et al., 2012). Maximum likelihood trees were configured and executed using RAxML (Stamatakis, 2014) with the GTR+GAMMA substitution model.

Bayesian phylogenetic analysis of full MTBC and L4 datasets

Bayesian phylogenetic analysis of the full MTBC was conducted using a dataset of 261 *M. tuberculosis* genomes including LUND1, five previously published ancient genomes (Bos et al., 2014; Kay et al., 2015), and 255 previously published modern genomes (Table

S4 in Additional File 1). *Mycobacterium canettii* was used as an outgroup for this dataset. Bayesian phylogenetic analysis of L4 of the MTBC was conducted using a dataset of 152 genomes including three ancient genomes presented here and in a previous publication (Kay et al., 2015) and 149 previously published modern genomes (Table S5 in Additional File 1). Body80 and body92 were selected out of the eight samples presented by Kay and colleagues based on multiple criteria. Multiple samples from that study proved to be mixed strain infections. Apart from body92, these samples were excluded from this analysis due to our present inability to separate strains without ignoring derived positions. Body92 had a clearly dominant strain estimated by Kay et al. (2015) to make up 96% of the tuberculosis data, and stringent mapping in BWA (Li and Durbin, 2009) (-l 32, -n 0.1, -q 37) found the genome to have 124-fold coverage when mapped against the TB ancestor. Between the degree of dominance and the high coverage, we could confidently call variant positions from the dominant strain (Figure S11a in Additional File 2). Body80 was the only single-strain sample from that collection to have sufficient coverage (~8x) for confident SNP calling after stringent mapping (Figure S11b in Additional File 2). For selection criteria for the modern genomes, please see Additional File 2. L2_N0020 was used as an outgroup. The possibility of equal evolutionary rates in both datasets was rejected by the MEGA-CC molecular clock test (Kumar et al., 2012). TempEst (Rambaut et al., 2016) was also used to assess temporal structure in the phylogeny prior to analysis with BEAST2 (Bouckaert et al., 2014). For the full MTBC alignment, $R^2=0.273$, and for the L4 alignment, $R^2=0.113$ (Figures S8 and S9 in Additional File 2). We generated a maximum likelihood tree and alignment excluding the of the full MTBC excluding the animal lineages, and consequently excluding the ancient *M. pinnipedii* genomes, to test if limiting the dataset to the human lineages produced a stronger temporal signal. Without the anchor of the ancient *M. pinnipedii* genomes, the temporal signal for the full complex reduced ($R^2=0.06$), as all ancient calibration points were limited to Lineage 4. When the root-to-tip distances are plotted with points labelled according to lineage or sublineage, it becomes clear that clade membership is largely driving the distance from root of the genomes. However, there remains a temporal signal in the data.

A correction for static positions in the *M. tuberculosis* genome not included in the SNP alignment was included in the configuration file. A “TVM” substitution model, selected based on results from ModelGenerator (Keane et al., 2006), was implemented in BEAUti as a GTR+G4 model with the AG rate parameter fixed to 1.0. LUND1, body80, and body92 were tip-calibrated using year of death, which was available for all three individuals (Table S5 in Additional File 1). The three ancient Peruvian genomes were calibrated using the mid-point of their OxCal ranges (Table S4 in Additional File 1) (Bos et al., 2014). We performed tip sampling for all modern genomes excluding the outgroup over a uniform distribution between 1992 and 2010 in all models. The outgroup was fixed to 2010 in every case. In the BDSKY models, all modern genomes were given a tip date of 2010. All tree priors were used in conjunction with an uncorrelated relaxed lognormal clock model. The constant coalescent model was also used in conjunction with a strict clock model.

Two independent MCMC chains of 200,000,000 iterations minimum were computed for each model. If the ESS for any parameter was below 200 after the chains were combined, they were resumed with additional iterations. The results were assessed in Tracer v1.7.1 with a 10 percent burn-in (Rambaut et al., 2014). Trees were sampled every 20,000 iterations. The log files and trees for each pair of runs were combined using LogCombiner v2.4.7 (Bouckaert et al., 2014). An MCC tree was generated using TreeAnnotator with ten percent burn-in (Bouckaert et al., 2014). For details on the parameterization of the birth-death models, please see Additional File 2.

Marginal likelihood was calculated using stepping stone sampling (Xie et al., 2011) implemented in the MODELSELECTION package in BEAST2. The total chain length required for convergence in each model was split across 100 steps. Following this, we performed a date randomization test (Ramsden et al., 2009) for the BDSKY+UCLD model for each dataset. Dates were shuffled randomly among all genomes excluding the outgroup. For both datasets the outgroup was used as an anchor for tip-dating of the “modern” genomes in each date-randomized model. Ten randomizations were generated for each model and run in at least two parallel chains until convergence.

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

Raw sequencing data from the non-UDG, non-enriched screening library and the UDG-treated, enriched library can be found under the BioProject PRJNA517266.

Competing interests

Not applicable

Funding

This project was supported by the Max Planck Society and by grants from the Erik Philip-Sörensen Foundation and Crafoord foundation to T.A. and C.A.

Authors' contributions

C.A. and K.I.B. conceived of the investigation. S.S., D.K., A.H., and K.I.B. designed the experiments. T.A., G.B., and C.A. performed the exhumation and radiological analysis of the mummy and provided a paleopathological examination. G.B. was responsible for the CT examinations together with imaging analysis and coordination of the calcification extraction. S.S. performed laboratory work. S.S., D.K., A.H., Å.J.V., and K.I.B. performed analyses.

Acknowledgments

The authors would like to thank Marta Burri for conducting the hybridization capture. We also thank Elizabeth A. Nelson for laboratory assistance; Maria Spyrou, Felix M. Key, and Luis Roger Esquivel Gomez for technical assistance; and all members of the Molecular Palaeopathology and Computational Pathogenomics research groups at the MPI-SHH for constructive comments throughout the study. We would like to acknowledge the Department of Imaging and Clinical Physiology, Skåne University Hospital Lund for the generous opportunity to examine the mummy of Peder Winstrup and the radiologists Roger Siemund,

Pär Wingren, Mats Geijer, Pernilla Gustavson and David Pellby for contributing to the image analyses. A sincere thank you to the thoracic surgeons Erik Gyllstedt and Jesper Andreasson for so delicately extracting the small calcification of interest and enabling further investigations.

REFERENCES

- Achtman, M., 2016. How old are bacterial pathogens? *Proc. R. Soc. B* 283, 20160990. <https://doi.org/10.1098/rspb.2016.0990>
- Andrades Valtueña, A., Mitnik, A., Key, F.M., Haak, W., Allmäe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., Massy, K., Novak, M., Pfrengle, S., Reinhold, S., Šlaus, M., Spyrou, M.A., Szécsényi-Nagy, A., Törv, M., Hansen, S., Bos, K.I., Stockhammer, P.W., Herbig, A., Krause, J., 2017. The Stone Age Plague and Its Persistence in Eurasia. *Current Biology*. <https://doi.org/10.1016/j.cub.2017.10.025>
- Armelagos, G.J., Brown, P.J., Turner, B., 2005. Evolutionary, historical and political economic perspectives on health and disease. *Social Science & Medicine, The Social Production of Health: Critical Contributions from Evolutionary, Biological and Cultural Anthropology: Papers in Memory of Arthur J. Rubel*The Social Production of Health: Critical Contributions from Evolutionary, Biological and Cultural Anthropology: Papers in Memory of Arthur J. Rubel 61, 755–765. <https://doi.org/10.1016/j.socscimed.2004.08.066>
- Auwer, G.A.V. der, Carneiro, M.O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* 43, 11.10.1-11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Bahr, N.C., Antinori, S., Wheat, L.J., Sarosi, G.A., 2015. Histoplasmosis infections worldwide: thinking outside of the Ohio River valley. *Curr Trop Med Rep* 2, 70–80. <https://doi.org/10.1007/s40475-015-0044-0>
- Baker, O., Lee, O.Y.-C., Wu, H.H.T., Besra, G.S., Minnikin, D.E., Llewellyn, G., Williams, C.M., Maixner, F., O'Sullivan, N., Zink, A., Chamel, B., Khawam, R., Coqueugniot, E., Helmer, D., Le Mort, F., Perrin, P., Gourichon, L., Dutailly, B., Pálfi, G., Coqueugniot, H., Dutour, O., 2015. Human tuberculosis predates domestication in ancient Syria. *Tuberculosis, Supplement issue: Tuberculosis in Evolution* 95, S4–S12. <https://doi.org/10.1016/j.tube.2015.02.001>
- Boritsch, E.C., Khanna, V., Pawlik, A., Honoré, N., Navas, V.H., Ma, L., Bouchier, C., Seemann, T., Supply, P., Stinear, T.P., Brosch, R., 2016. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. *PNAS* 201604921. <https://doi.org/10.1073/pnas.1604921113>
- Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514, 494–497. <https://doi.org/10.1038/nature13591>
- Boskova, V., Stadler, T., Magnus, C., 2018. The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. *Virus Evol* 4. <https://doi.org/10.1093/ve/vex044>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* 10, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., Pääbo, S., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38, e87–e87. <https://doi.org/10.1093/nar/gkp1163>
- Brown, K., Mund, D.F., Aberle, D.R., Batra, P., Young, D.A., 1994. Intrathoracic calcifications: radiographic features and differential diagnoses. *RadioGraphics* 14, 1247–1261. <https://doi.org/10.1148/radiographics.14.6.7855339>

- Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J., Alfsnes, K., Pettersson, J.O.-H., Kirkeleite, I., Fandinho, F., Silva, M.A. da, Perdigao, J., Portugal, I., Viveiros, M., Clark, T., Caws, M., Dunstan, S., Thai, P.V.K., Lopez, B., Ritacco, V., Kitchen, A., Brown, T.S., Soolingen, D. van, O'Neill, M.B., Holt, K.E., Feil, E.J., Mathema, B., Balloux, F., Eldholm, V., 2018. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Science Advances* 4, eaat5869. <https://doi.org/10.1126/sciadv.aat5869>
- Burrill, J., Williams, C.J., Bain, G., Conder, G., Hine, A.L., Misra, R.R., 2007. Tuberculosis: a radiologic review. *Radiographics* 27, 1255–1273. <https://doi.org/10.1148/rg.275065176>
- Canci, A., Minozzi, S., Tarli, S.M.B., 1996. New Evidence of Tuberculous Spondylitis from Neolithic Liguria (Italy). *International Journal of Osteoarchaeology* 6, 497–501. [https://doi.org/10.1002/\(SICI\)1099-1212\(199612\)6:5<497::AID-OA291>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1099-1212(199612)6:5<497::AID-OA291>3.0.CO;2-O)
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>
- Cohen, M.N., Armelagos, G.J., 1984. *Paleopathology at the Origins of Agriculture*. University Press of Florida, Gainesville, Florida.
- Coll, F., McNerney, R., Guerra-Assunção, J.A., Glynn, J.R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nature Communications* 5, 4812. <https://doi.org/10.1038/ncomms5812>
- Comas, I., Chakravarti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., Gagneux, S., 2010. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 42, 498–503. <https://doi.org/10.1038/ng.590>
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 45, 1176–1182. <https://doi.org/10.1038/ng.2744>
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 110, 15758–15763. <https://doi.org/10.1073/pnas.1314445110>
- Dabney, Jesse, Meyer, M., Pääbo, S., 2013. Ancient DNA Damage. *Cold Spring Harb Perspect Biol* 5, a012567. <https://doi.org/10.1101/cshperspect.a012567>
- Didelot, X., Wilson, D.J., 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology* 11, e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed Phylogenetics and Dating with Confidence. *PLOS Biology* 4, e88. <https://doi.org/10.1371/journal.pbio.0040088>
- Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. *Trends in Ecology & Evolution* 18, 481–488. [https://doi.org/10.1016/S0169-5347\(03\)00216-7](https://doi.org/10.1016/S0169-5347(03)00216-7)
- Dubos, R., Dubos, J., 1952. *The White Plague: Tuberculosis, Man, and Society*, 3rd ed. Rutgers University Press, New Brunswick, New Jersey.
- Duchêne, S., Duchêne, D., Holmes, E.C., Ho, S.Y.W., 2015. The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data. *Mol Biol Evol* 32, 1895–1906. <https://doi.org/10.1093/molbev/msv056>
- Duchêne, S., Holmes, E.C., Ho, S.Y.W., 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc. B* 281, 20140732. <https://doi.org/10.1098/rspb.2014.0732>

- El-Najjar, M., Al-Shiyab, A., Al-Sarie, I., 1996. Cases of tuberculosis at 'Ain Ghazal, Jordan. *Paléorient* 22, 123–128. <https://doi.org/10.3406/paleo.1996.4639>
- Ford, C.B., Lin, P.L., Chase, M.R., Shah, R.R., Iartchouk, O., Galagan, J., Mohaideen, N., Iøerger, T.R., Sacchettini, J.C., Lipsitch, M., Flynn, J.L., Fortune, S.M., 2011. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nature Genetics* 43, 482–486. <https://doi.org/10.1038/ng.811>
- Formicola, V., Milanese, Q., Scarsini, C., 1987. Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *American Journal of Physical Anthropology* 72, 1–6. <https://doi.org/10.1002/ajpa.1330720102>
- Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., Pääbo, S., 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *PNAS* 110, 2223–2227. <https://doi.org/10.1073/pnas.1221359110>
- Gagneux, S., 2018. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology* 16, 202–213. <https://doi.org/10.1038/nrmicro.2018.8>
- Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., Jong, B.C. de, Narayanan, S., Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C., Hilty, M., Hopewell, P.C., Small, P.M., 2006. Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *PNAS* 103, 2869–2873. <https://doi.org/10.1073/pnas.0511240103>
- Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., Meyer, M., 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res* 45, e79–e79. <https://doi.org/10.1093/nar/gkx033>
- Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E., Orlando, L., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27, 2153–2155. <https://doi.org/10.1093/bioinformatics/btr347>
- Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., Supply, P., Vincent, V., 2005. Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. *PLOS Pathogens* 1, e5. <https://doi.org/10.1371/journal.ppat.0010005>
- Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W., Gagneux, S., 2008. High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography. *PLOS Biology* 6, e311. <https://doi.org/10.1371/journal.pbio.0060311>
- Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y.-C., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., Bar-Gal, G.K., Spigelman, M., 2008. Detection and Molecular Characterization of 9000-Year-Old *Mycobacterium tuberculosis* from a Neolithic Settlement in the Eastern Mediterranean. *PLoS ONE* 3, e3426. <https://doi.org/10.1371/journal.pone.0003426>
- Ho, S.Y.W., Larson, G., 2006. Molecular clocks: when times are a-changin'. *Trends in Genetics* 22, 79–83. <https://doi.org/10.1016/j.tig.2005.11.006>
- Ho, S.Y.W., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Mol Biol Evol* 22, 1561–1568. <https://doi.org/10.1093/molbev/msi145>
- Ho, S.Y.W., Shapiro, B., Phillips, M.J., Cooper, A., Drummond, A.J., 2007. Evidence for Time Dependency of Molecular Rate Estimates. *Syst Biol* 56, 515–522. <https://doi.org/10.1080/10635150701435401>
- Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., Hannon, G.J., 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat. Protocols* 4, 960–974. <https://doi.org/10.1038/nprot.2009.68>
- Houben, R.M.G.J., Dodd, P.J., 2016. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Medicine* 13, e1002152. <https://doi.org/10.1371/journal.pmed.1002152>

- Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* 12, e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>
- Jankute, M., Nataraj, V., Lee, O.Y.-C., Wu, H.H.T., Ridell, M., Garton, N.J., Barer, M.R., Minnikin, D.E., Bhatt, A., Besra, G.S., 2017. The role of hydrophobicity in tuberculosis evolution and pathogenicity. *Scientific Reports* 7, 1315. <https://doi.org/10.1038/s41598-017-01501-0>
- Kay, G.L., Sergeant, M.J., Zhou, Z., Chan, J.Z.-M., Millard, A., Quick, J., Szikossy, I., Pap, I., Spigelman, M., Loman, N.J., Achtman, M., Donoghue, H.D., Pallen, M.J., 2015. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nature Communications* 6, 6717. <https://doi.org/10.1038/ncomms7717>
- Keane, T., Creevey, C., Pentony, M., Naughton, T.J., McInerney, J., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology* 6, 29–46.
- Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* 40, e3–e3. <https://doi.org/10.1093/nar/gkr771>
- Köhler, K., Pálfi, G., Molnár, E., Zalai-Gaál, I., Oszás, A., Bánffy, E., Kirinó, K., Kiss, K.K., Mende, B.G., 2014. A Late Neolithic Case of Pott's Disease from Hungary. *International Journal of Osteoarchaeology* 24, 697–703. <https://doi.org/10.1002/oa.2254>
- Kumar, S., Stecher, G., Peterson, D., Tamura, K., 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 28, 2685–2686. <https://doi.org/10.1093/bioinformatics/bts507>
- Leung, A.N., Müller, N.L., Pineda, P.R., FitzGerald, J.M., 1992. Primary tuberculosis in childhood: radiographic manifestations. *Radiology* 182, 87–91. <https://doi.org/10.1148/radiology.182.1.1727316>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Loosdrecht, M. van de, Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., Hajraoui, M.A.E., Amzazi, S., Hublin, J.-J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W., Jeong, C., Krause, J., 2018. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* 360, 548–552. <https://doi.org/10.1126/science.aar8380>
- Mann, A.E., Sabin, S., Ziesemer, K., Vågane, Å.J., Schroeder, H., Ozga, A.T., Sankaranarayanan, K., Hofman, C.A., Yates, J.A.F., Salazar-García, D.C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G.R., Stone, A.C., Lewis, C.M., Krause, J., Hofman, C., Bos, K.I., Warinner, C., 2018. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific Reports* 8, 9822. <https://doi.org/10.1038/s41598-018-28091-9>
- Masson, M., Molnár, E., Donoghue, H.D., Besra, G.S., Minnikin, D.E., Wu, H.H.T., Lee, O.Y.-C., Bull, I.D., Pálfi, G., 2013. Osteological and Biomolecular Evidence of a 7000-Year-Old Case of Hypertrophic Pulmonary Osteopathy Secondary to Tuberculosis from Neolithic Hungary. *PLOS ONE* 8, e78252.
- Menardo, F., Duchene, S., Brites, D., Gagneux, S., 2019. The molecular clock of *Mycobacterium tuberculosis*. *bioRxiv* 532390. <https://doi.org/10.1101/532390>
- Meyer, M., Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* 2010, pdb.prot5448-pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Möller, S., Plessis, L. du, Stadler, T., 2018. Impact of the tree prior on estimating clock rates during epidemic outbreaks. *PNAS* 201713314. <https://doi.org/10.1073/pnas.1713314115>

- Neukamm, J., Peltzer, A., 2018. Integrative-Transcriptomics/DamageProfiler.
- Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution. *PNAS* 96, 12638–12643. <https://doi.org/10.1073/pnas.96.22.12638>
- O'Neill, M.B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2019. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *Mol Ecol* 15120. <https://doi.org/10.1111/mec.15120>
- O'Neill, M.B., Shockey, A.C., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2018. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *bioRxiv* 210161. <https://doi.org/10.1101/210161>
- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Knip, C., Krause, J., Nieselt, K., 2016. EAGER: efficient ancient genome reconstruction. *Genome Biology* 17, 60. <https://doi.org/10.1186/s13059-016-0918-z>
- Pepperell, C.S., Casto, A.M., Kitchen, A., Granka, J.M., Comejo, O.E., Holmes, E.C., Birren, B., Galagan, J., Feldman, M.W., 2013. The Role of Selection in Shaping Diversity of Natural *M. tuberculosis* Populations. *PLOS Pathogens* 9, e1003543. <https://doi.org/10.1371/journal.ppat.1003543>
- Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2. <https://doi.org/10.1093/ve/vew007>
- Rambaut, A., Suchard, M., Xie, W., Drummond, A.J., 2014. Tracer. Institute of Evolutionary Biology, University of Edinburgh.
- Ramsden, C., Holmes, E.C., Charleston, M.A., 2009. Hantavirus Evolution in Relation to Its Rodent and Insectivore Hosts: No Evidence for Codivergence. *Mol Biol Evol* 26, 143–153. <https://doi.org/10.1093/molbev/msn234>
- Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., Nielsen, H.B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M.V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Pontén, T., Foley, R.A., Lahr, M.M., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* 163, 571–582. <https://doi.org/10.1016/j.cell.2015.10.009>
- Renaud, G., Slon, V., Duggan, A.T., Kelso, J., 2015. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology* 16, 1–18. <https://doi.org/10.1186/s13059-015-0776-0>
- Roberts, C.A., Buikstra, J.E., 2003. *The Bioarchaeology of Tuberculosis: A Global View on a Reemerging Disease*. University Press of Florida, Gainesville, Florida.
- Rothschild, B.M., Martin, L., Lev, G., Bercovier, H., Bar-Gal, G.K., Greenblatt, C.L., Donoghue, H.D., Spigelman, M., Brittain, D., 2001. *Mycobacterium tuberculosis* Complex DNA from an Extinct Bison Dated 17,000 Years before the Present. *Clinical Infectious Diseases* 305–311.
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., Boucher, C.A., 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415, 497–502. <https://doi.org/10.1038/415497a>
- Schuenemann, V.J., Peltzer, A., Welte, B., van Pelt, W.P., Molak, M., Wang, C.-C., Furtwängler, A., Urban, C., Reiter, E., Nieselt, K., Teßmann, B., Francken, M., Harvati, K., Haak, W., Schiffels, S., Krause, J., 2017. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications* 8, 15694. <https://doi.org/10.1038/ncomms15694>
- Shillington, K., 2012. *History of Africa*, Third. ed. Palgrave MacMillan, New York.
- Sparacello, V.S., Roberts, C.A., Kerudin, A., Müller, R., 2017. A 6500-year-old Middle Neolithic child from Pollera Cave (Liguria, Italy) with probable multifocal osteoarticular

- tuberculosis. *International Journal of Paleopathology*.
<https://doi.org/10.1016/j.ijpp.2017.01.004>
- Spyrou, M.A., Tukhbatova, R.I., Wang, C.-C., Valtueña, A.A., Lankapalli, A.K., Kondrashin, V.V., Tsybin, V.A., Khokhlov, A., Kühnert, D., Herbig, A., Bos, K.I., Krause, J., 2018. Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nature Communications* 9, 2234. <https://doi.org/10.1038/s41467-018-04550-9>
- Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *PNAS* 94, 9869–9874.
- Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A.J., 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *PNAS* 110, 228–233. <https://doi.org/10.1073/pnas.1207965110>
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
<https://doi.org/10.1093/bioinformatics/btu033>
- Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., Fenner, L., Rutaiwa, L., Borrell, S., Luo, T., Gao, Q., Kato-Maeda, M., Ballif, M., Egger, M., Macedo, R., Mardassi, H., Moreno, M., Vilanova, G.T., Fyfe, J., Globan, M., Thomas, J., Jamieson, F., Guthrie, J.L., Asante-Poku, A., Yeboah-Manu, D., Wampande, E., Ssengooba, W., Joloba, M., Boom, W.H., Basu, I., Bower, J., Saraiva, M., Vasconcellos, S.E.G., Suffys, P., Koch, A., Wilkinson, R., Gail-Bekker, L., Malla, B., Ley, S.D., Beck, H.-P., de Jong, B.C., Toit, K., Sanchez-Padilla, E., Bonnet, M., Gil-Brusola, A., Frank, M., Penlap Beng, V.N., Eisenach, K., Alani, I., Ndung'u, P.W., Revathi, G., Gehre, F., Akter, S., Ntoumi, F., Stewart-Isherwood, L., Ntinginya, N.E., Rachow, A., Hoelscher, M., Cirillo, D.M., Skenders, G., Hoffner, S., Bakonyte, D., Stakenas, P., Diel, R., Crudu, V., Moldovan, O., Al-Hajoj, S., Otero, L., Barletta, F., Carter, E.J., Diero, L., Supply, P., Comas, I., Niemann, S., Gagneux, S., 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* 48, 1535–1543.
<https://doi.org/10.1038/ng.3704>
- Vågene, Å.J., Herbig, A., Campana, M.G., García, N.M.R., Warinner, C., Sabin, S., Spyrou, M.A., Valtueña, A.A., Huson, D., Tuross, N., Bos, K.I., Krause, J., 2018. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution* 520–528. <https://doi.org/10.1038/s41559-017-0446-6>
- Warinner, C., Herbig, A., Mann, A., Yates, J.A.F., Weiß, C.L., Burbano, H.A., Orlando, L., Krause, J., 2017. A Robust Framework for Microbial Archaeology. *Annual Review of Genomics and Human Genetics* 18, null. <https://doi.org/10.1146/annurev-genom-091416-035526>
- WHO, 2018. WHO | Tuberculosis (TB) [WWW Document]. WHO. URL <http://www.who.int/tb/en/> (accessed 11.18.18).
- Wilbur, A.K., Bouwman, A.S., Stone, A.C., Roberts, C.A., Pfister, L.-A., Buikstra, J.E., Brown, T.A., 2009. Deficiencies and challenges in the study of ancient tuberculosis DNA. *Journal of Archaeological Science* 36, 1990–1997.
<https://doi.org/10.1016/j.jas.2009.05.020>
- Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., Soolingen, D. van, Rüsç-Gerdes, S., Locht, C., Brisse, S., Meyer, A., Supply, P., Niemann, S., 2008. Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. *PLOS Pathogens* 4, e1000160. <https://doi.org/10.1371/journal.ppat.1000160>
- Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.-H., 2011. Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection. *Syst Biol* 60, 150–160.
<https://doi.org/10.1093/sysbio/syq085>

ADDITIONAL FILES

Additional File 1

Format: Excel spreadsheet (.xlsx)

Title: Supplementary Tables

Description: Large tables of data contributing to the analyses presented in this paper, including a taxon table showing assigned reads from all taxonomic levels represented in the metagenomic LUND1 library (Table S1); full EAGER pipeline results for LUND1 shotgun sequencing data when mapped to HG19 human reference genome and TB ancestor genome, the non-UDG-treated enriched LUND1 data when mapped to the TB ancestor genome, and the UDG-treated enriched LUND1 data when mapped to the TB ancestor genome (Table S2); full EAGER pipeline results for negative controls processed with LUND1, mapped to the reconstructed TB ancestor genome (Table S3); genomes included in the full MTBC dataset, with respective publications, accession numbers, lineages, and dates (when applicable) (Table S4); genomes included in the L4 dataset, with respective publications, accession numbers, lineages, dates, and percentage of total SNPs called as heterozygous (Table S5); sites excluded from the full MTBC dataset (Table S6); sites excluded from the L4 dataset (Table S7); SnpEff annotation for derived alleles in LUND1 (Table S8).

Additional File 2

Format: Word document (.docx)

Title: Supplementary Information

Description: Detailed supplements to the RESULTS and METHODS sections, including supplementary figures

ADDITIONAL FILE 2

SUPPLEMENTARY INFORMATION

For

A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex.

Mitochondrial analysis

Given the uncharacteristically high endogenous human DNA content of the lung nodule, we estimated the mitochondrial contamination rate using Schmutzi (Renaud et al., 2015). In order to do this we mapped the metagenomic sequencing data to the hg19 mitochondrial chromosome with CircularMapper as implemented in the EAGER pipeline (Peltzer et al., 2016). We called initial contamination estimates with contDeam, and then ran Schmutzi, calling a consensus mitochondrial sequence and making a final average contamination estimate of 2% (lower boundary 1%, upper boundary 3%).

Heterozygosity analysis and selection of modern L4 genomes

We chose to reduce the quantity of L4 genomes used for Bayesian phylogeny in this study for the purposes of computational feasibility and attempting to balance the representation of different L4 sublineages without compromising allelic diversity in the analysis. The four deeply sampled sublineages published by Stucki and colleagues (2016) – L4.3/LAM, L4.6.1/Uganda, L4.1.2/Haarlem, and L4.10/PGG3 – were subsampled to 30 representatives each based on the quantity of heterozygous positions found in each genome after the inclusion of strains published in the Comas et al. (2013, 2010) datasets. For each of these four sublineages, the 30 genomes with the fewest heterozygous positions were selected for inclusion. We quantified the number of heterozygous positions using MultiVCFAnalyzer (v0.87 <https://github.com/alexherbig/MultiVCFAnalyzer>) (Bos et al., 2014) (see METHODS).

SNP Effect Analysis

SnEff (v4.2) (Cingolani et al., 2012) was run for LUND1 using a gene annotation database tailored to the TB ancestor reference genome (Comas et al., 2010). With the exception of setting the upstream and downstream effect interval to 100 nucleotides, we used default settings. Variant positions unique to LUND1 in comparison to the L4 dataset alignment were filtered, and the SnEff results for these positions, in addition to gene descriptions from the NCBI Gene database, can be found in Table S9 of Additional File 1.

Of note were two mce-associated genes thought to be involved in macrophage invasion (Gagneux, 2018) that showed MODERATE or MODIFIER impact annotations: *mce1C* (missense variant) and *mce1B* (downstream gene variant). A MODERATE impact variant (missense variant) was also flagged in *zmp1*, which may be involved in host-pathogen interaction (Correa et al., 2014).

Additionally, there were numerous variant positions that had MODERATE or MODIFIER impact annotations in genes involved in cell wall or membrane functions, as determined through Mycobrowser (Kapopoulou et al., 2011): *mmpL4* (missense variant), *mmpS4* (downstream gene variant), Rv0102 (upstream gene variant), Rv0912 (downstream gene variant), *ctpG* (missense variant), Rv1999c (missense variant), *cysA1* (missense variant), *cysW* (missense variant), *lppR* (downstream gene variant), *merT* (missense variant), Rv3104c (downstream gene variant), Rv3273 (missense variant), Rv3635 (downstream gene variant), and Rv3821 (downstream gene variant).

Birth-death model parameterization

For both datasets, birth death multi-rho tree priors were utilized. The L4 birth death skyline analyses were parameterized as follows for both BDSKY+UCLD+origin and BDSKY+UCLD. The rho parameter, referencing the sampling proportion at each sampling time, was split into four dimensions. One dimension was provided for each ancient sample, with a Beta prior distribution with mean 0.01, and one dimension was provided for modern genome representation, with Beta prior distribution with mean 0.1. The rho sampling times for each ancient genome were given as their respective tip dates, and the rho sampling time for modern genomes was set to 0. The reproductive number parameter (R), was given 5

dimensions, which provided estimates of population dynamics within L4 over time. The `becomeUninfectiousRate` was given one dimension for the ancient data with a Beta prior distribution with mean 0.001, and one dimension for the modern data with a Beta prior distribution of 0.1. For the L4 BDSKY+UCLD+origin analysis we imposed an upper limit on the origin of 10,000 years, with a starting value of 4,000 years (the initial value of the origin must be greater than the initial tree height estimate), and a uniform distribution.

MTBC birth-death analysis (BDSKY+UCLD) was parameterized as follows. The ρ parameter was split into seven dimensions with one for each ancient genome, with a Beta prior distribution centered around 0.001 for dimensions one to six, and 0.1 for the seventh dimension, translating to 1% and 10% sampling probabilities for ancient and modern genomes, respectively. The sampling times for ρ were set as the tip date for each ancient genome, and 0 for modern genomes. R was set to five dimensions. The `becomeUninfectiousRate` was given one dimension for the ancient data with a Beta prior distribution with mean 0.001, and one dimension for the modern data with a Beta prior distribution of 0.1.

Reproductive number (R) estimates

In an outbreak setting, R refers to the average number of secondary cases stemming from a single infection, and an epidemic event is inferred when the value is greater than one. However, for the data at hand, $R > 1$ translates to lineage diversification rates exceeding lineage death/extinction. A thorough phylodynamic analysis using the birth-death skyline population model requires inclusion of “outbreak” samples only. This shall be explored in future work. The R estimates below can only be tentatively interpreted.

In the full MTBC dataset, around 1300 BP the 95% HPD excludes 1, indicating a positive net diversification rate, with a significant increase between 974 and 390 BP (odds ratio=10.00054) (Figure S14). For both datasets, we see an increase in R at approximately 750 BP. In the MTBC model, it increases sharply and maintains its peak between 4 and 5. The increase is more gradual in the L4 model (Figure S15), and declines to hovering just above $R=1$. This roughly coincides with a jump in effective population size estimated by Liu

and colleagues for MTBC lineages indigenous to China (Liu et al., 2018). The decline of R for L4 beginning approximately 350 BP appears surprising, given the historically recorded rise of the White Plague in Europe from the 17th-19th century (Dubos and Dubos, 1952).

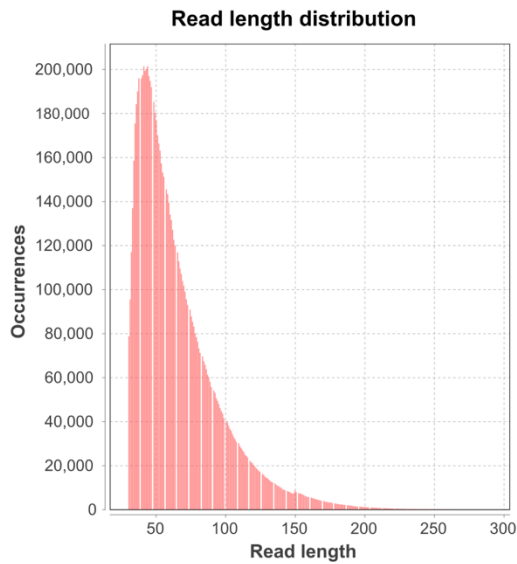


Figure S1. Fragment length distribution plot for 150 cycle, paired-end sequencing data from UDG-treated, TB captured LUND1 library. Fragment lengths as provided by DamageProfiler in the EAGER pipeline for 9,482,901 reads mapped to the TB ancestor genome with BWA as implemented in EAGER (-l 32, -n 0.1, -q 37).



Figure S2. MTBC dataset neighbor joining tree. The neighbor joining tree was configured using MEGA-Proto and generated with MEGA-CC, with 500 bootstrap replicates. The red stars indicate the ancient genomes.

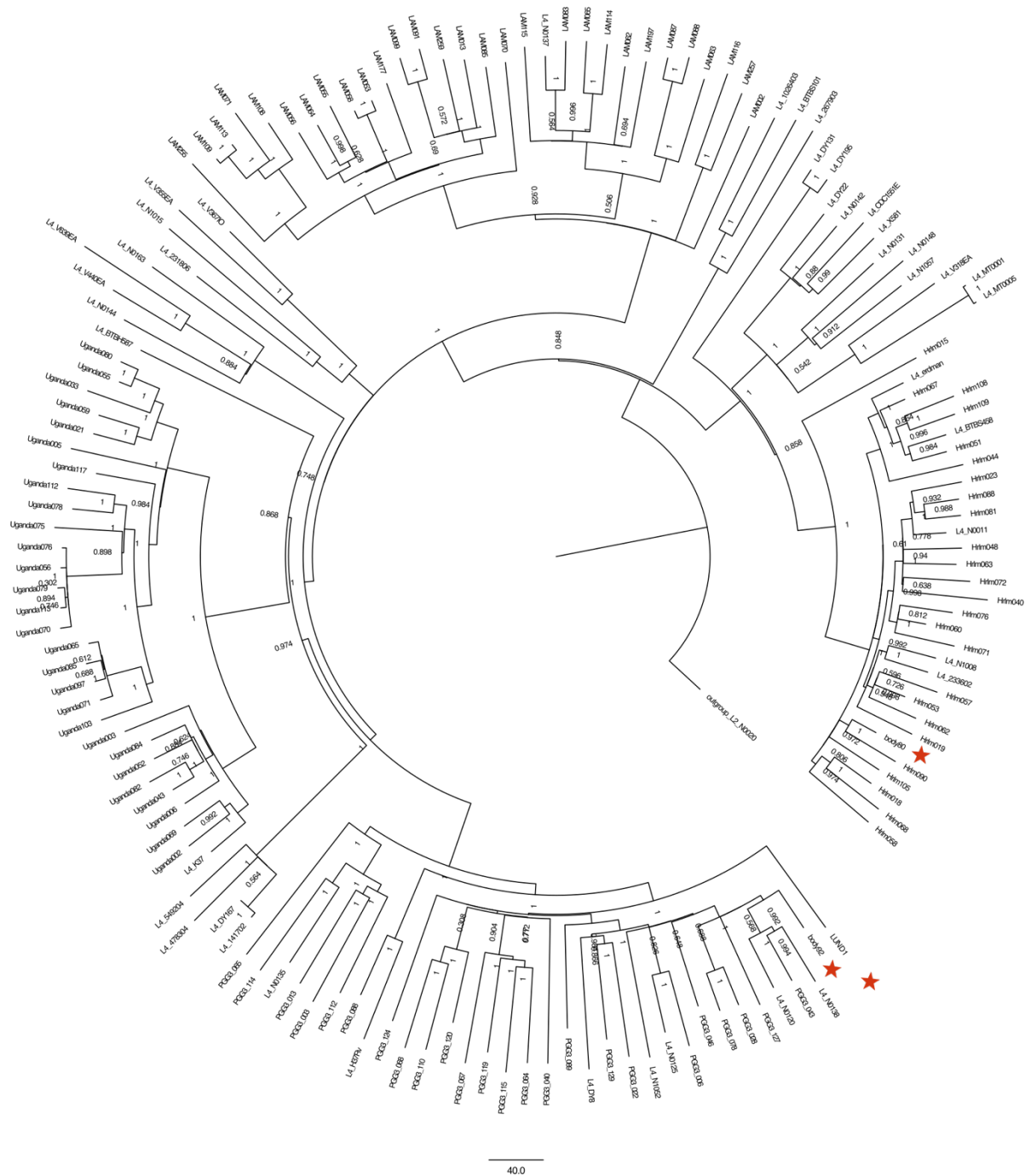


Figure S3. L4 dataset neighbor joining tree. The neighbor joining tree was configured using MEGA-Proto and generated with MEGA-CC, with 500 bootstrap replicates. The red stars indicate the ancient genomes.

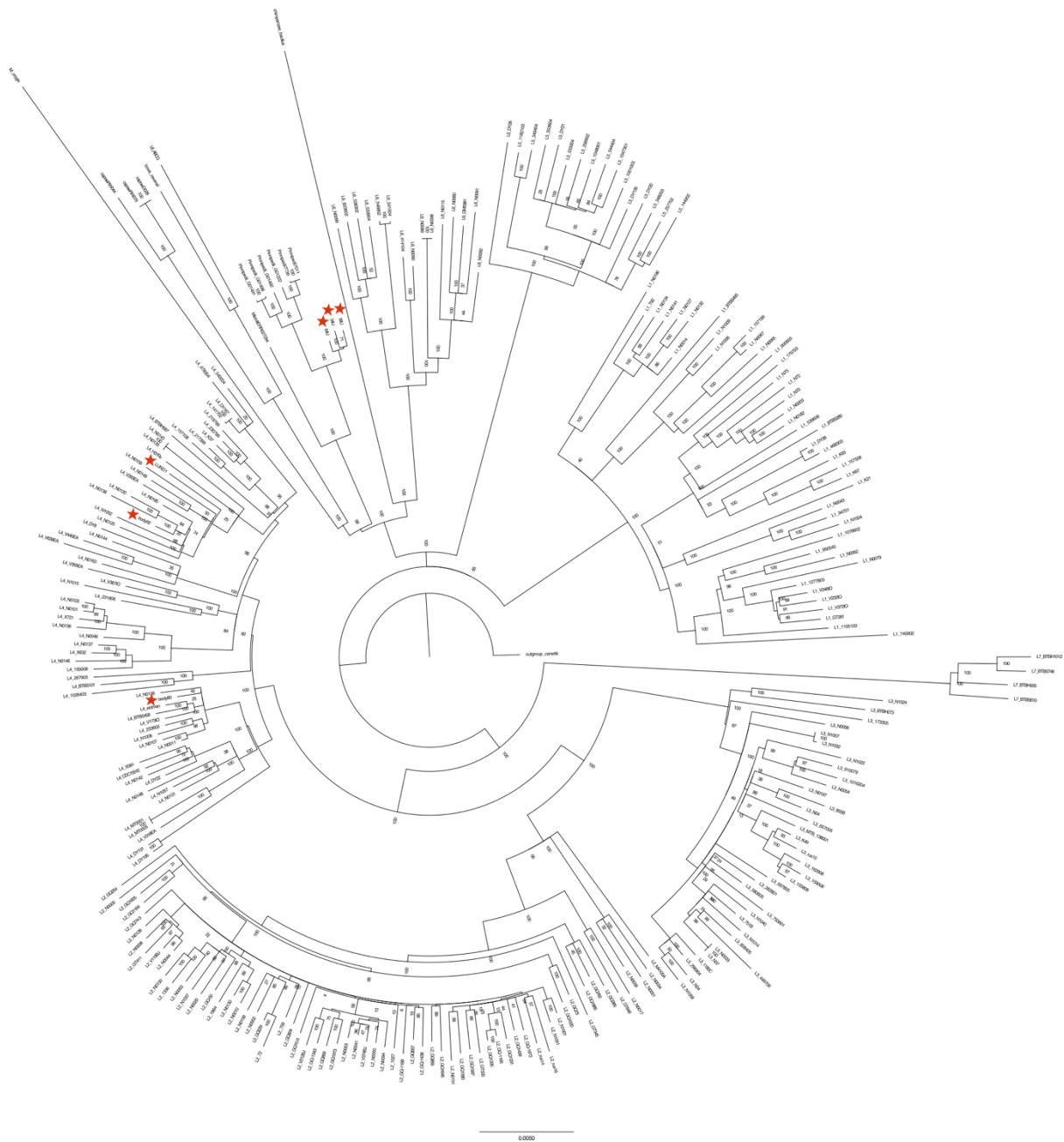


Figure S4. MTBC dataset maximum likelihood tree. The maximum likelihood tree was configured and generated with RAxML, with 500 bootstrap replicates. The red stars indicate the ancient genomes.

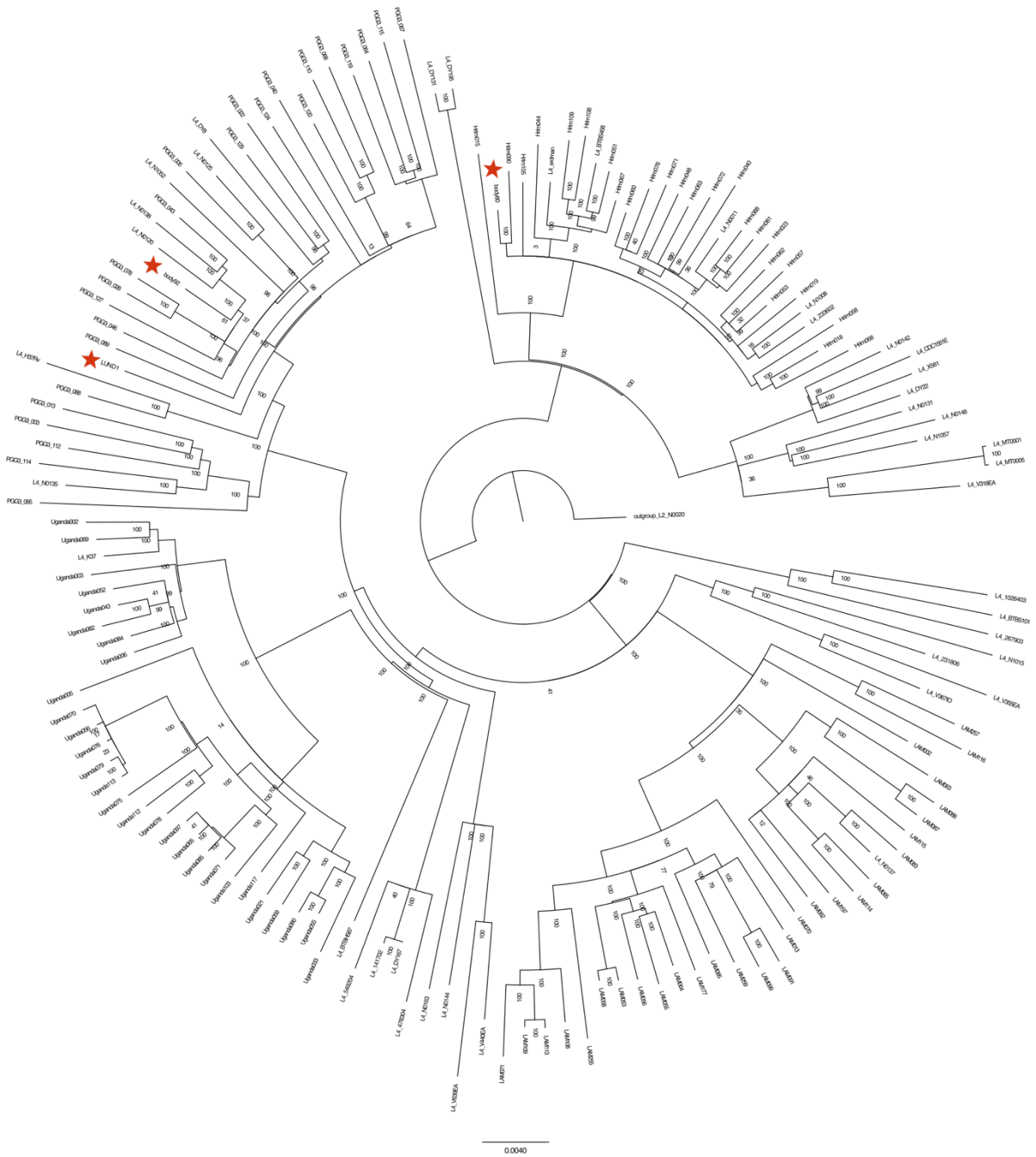


Figure S5. L4 dataset maximum likelihood tree. The maximum likelihood tree was configured and generated with RAXML, with 500 bootstrap replicates. The red stars indicate the ancient genomes.

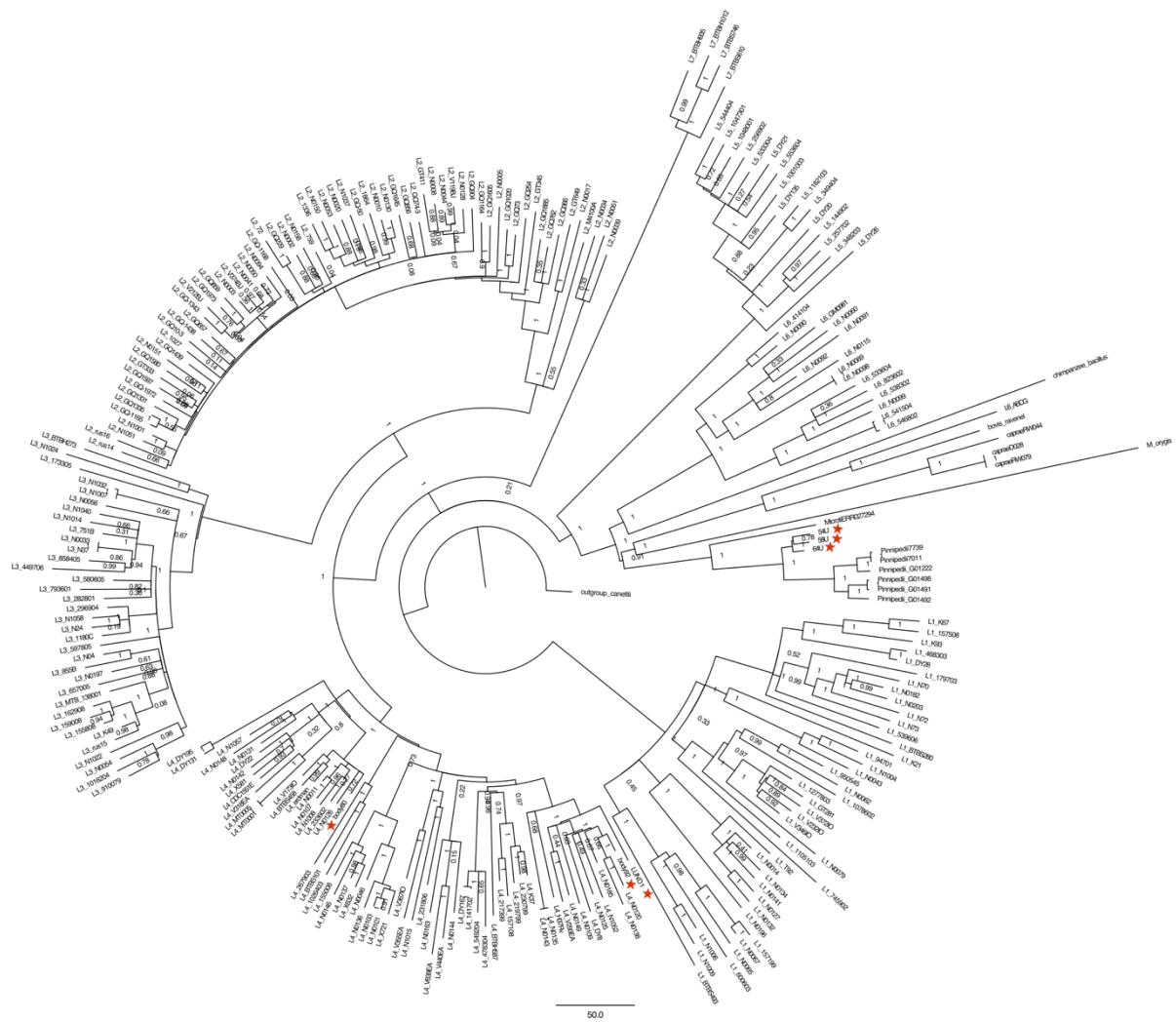


Figure S6. MTBC dataset maximum parsimony tree. The maximum parsimony tree was configured using MEGA-Proto and generated with MEGA-CC, with 500 bootstrap replicates. The red stars indicate the ancient genomes.

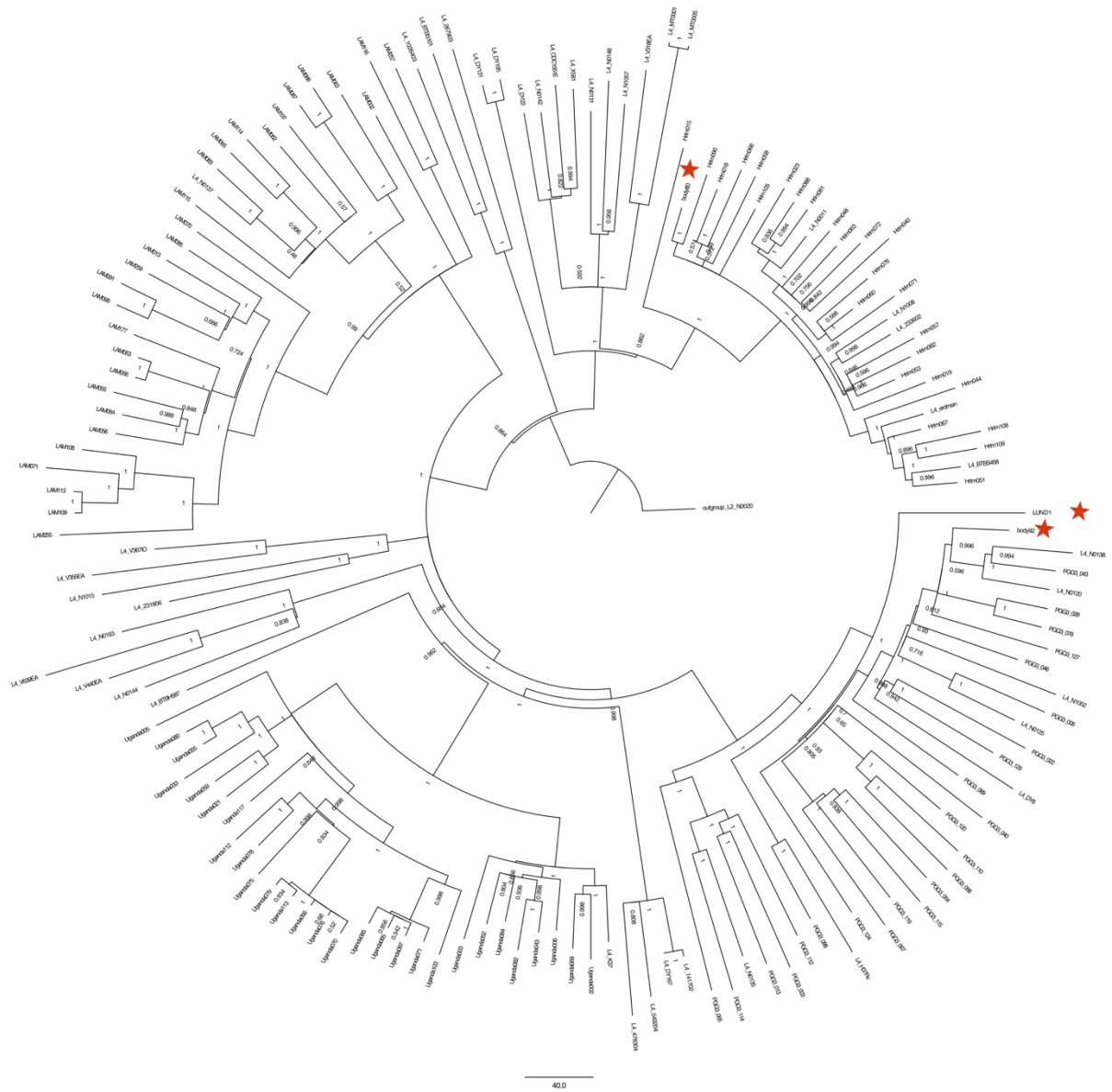


Figure S7. L4 dataset maximum parsimony tree. The maximum parsimony tree was configured using MEGA-Proto and generated with MEGA-CC, with 500 bootstrap replicates. The red stars indicate the ancient genomes.

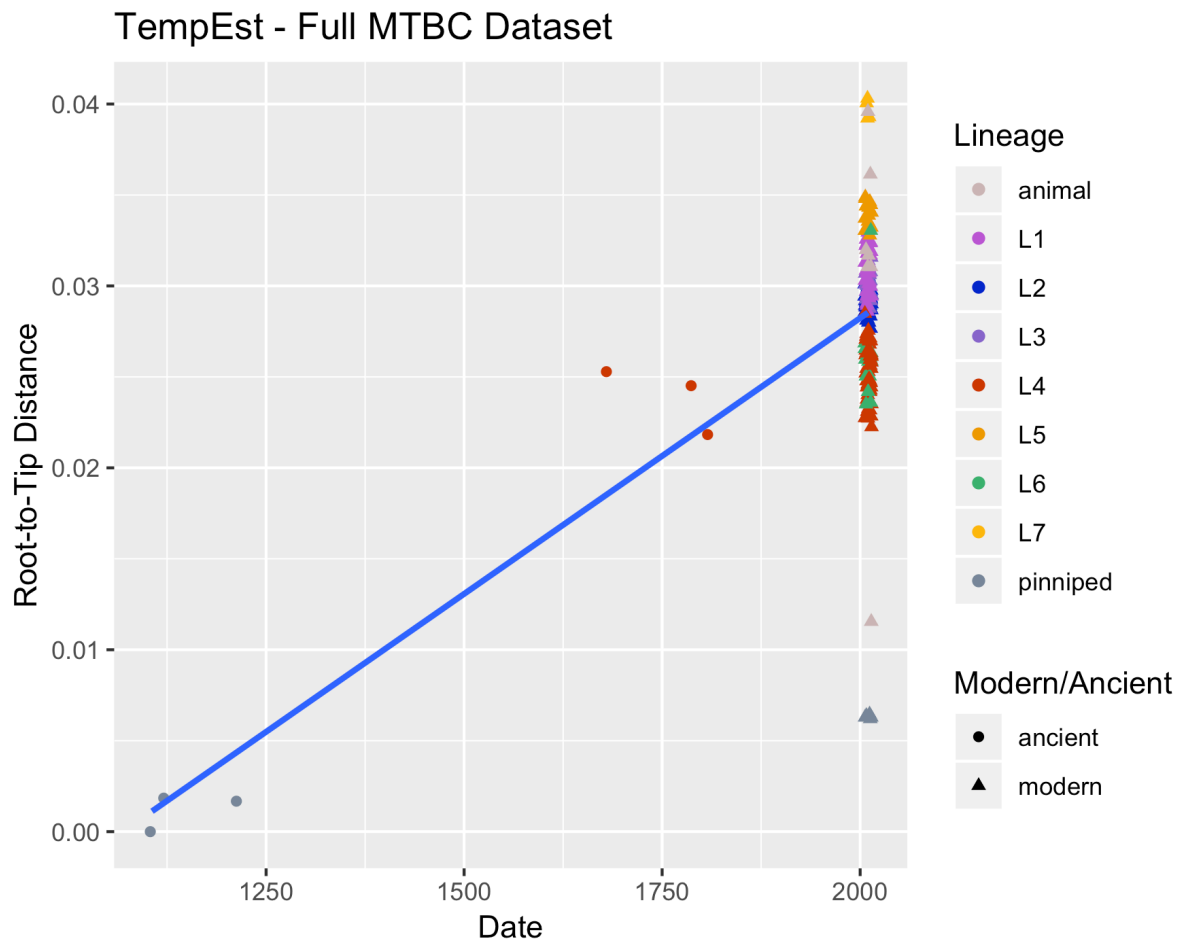


Figure S8. TempEst plot for full MTBC dataset. Plot generated from root-to-tip distances calculated in TempEst. $R^2 = 0.27$.

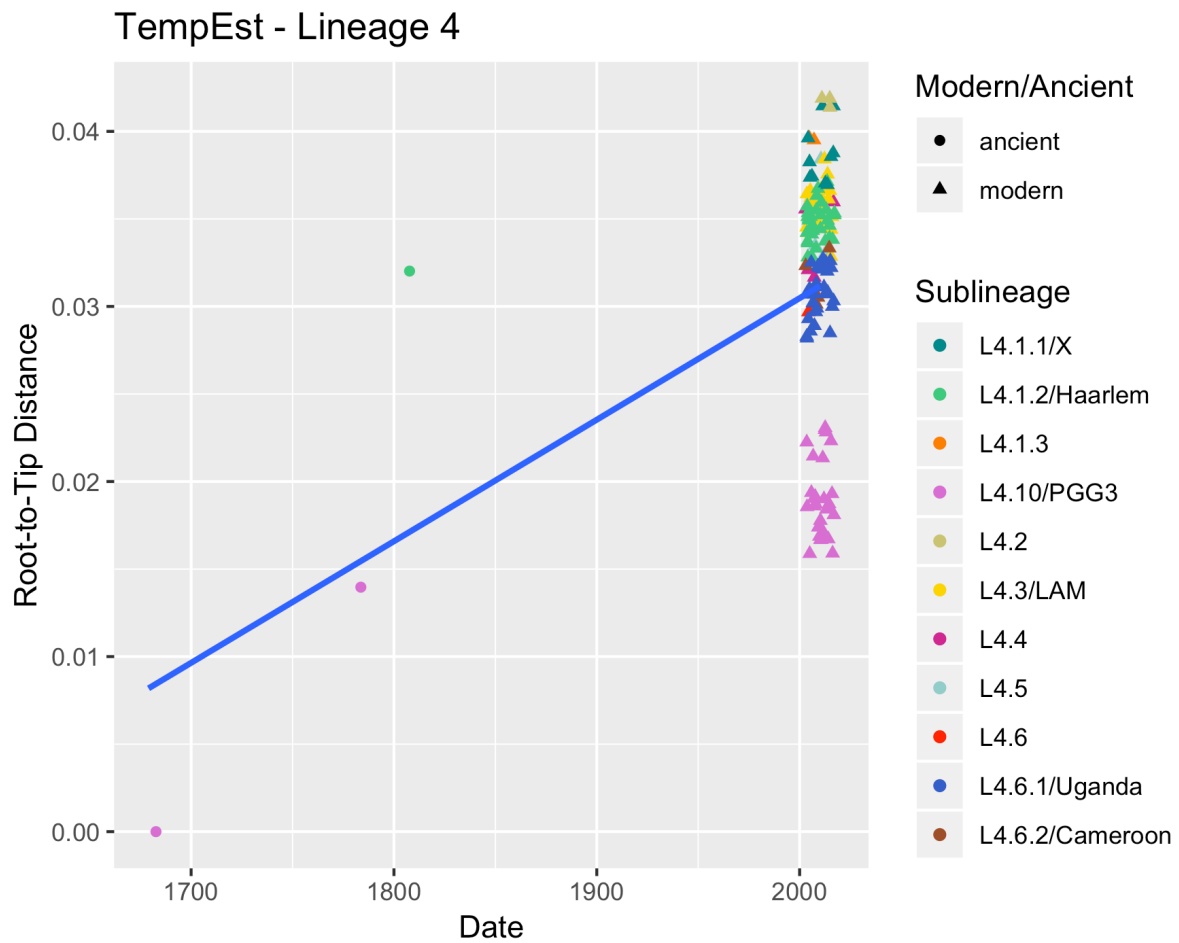


Figure S9. TempEst plot for Lineage 4 dataset. Plot generated from root-to-tip distances calculated in TempEst. $R^2 = 0.11$.

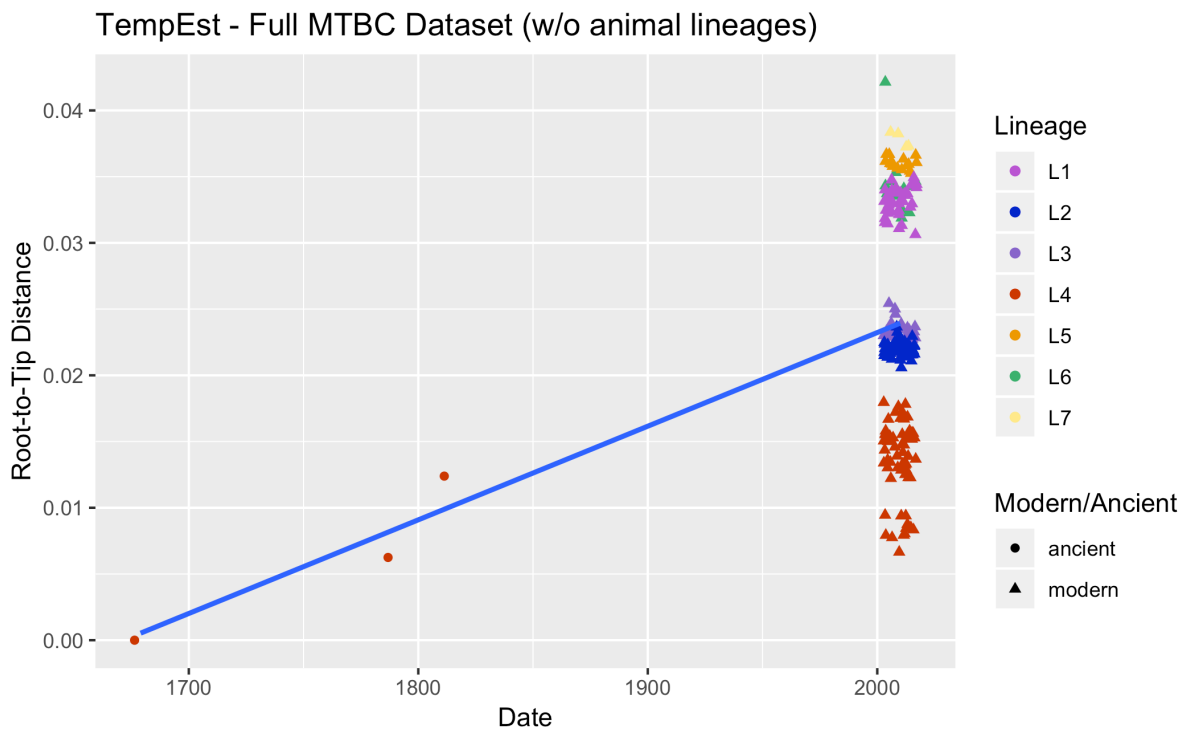


Figure S10. TempEst plot for full MTBC excluding animal lineages. Plot generated from root-to-tip distances calculated in TempEst. Input was a maximum likelihood tree generated with RAxML based on a SNP alignment excluding animal lineages of the MTBC, and consequently excluding the ancient *M. pinnipedii* genomes. $R^2 = 0.06$.

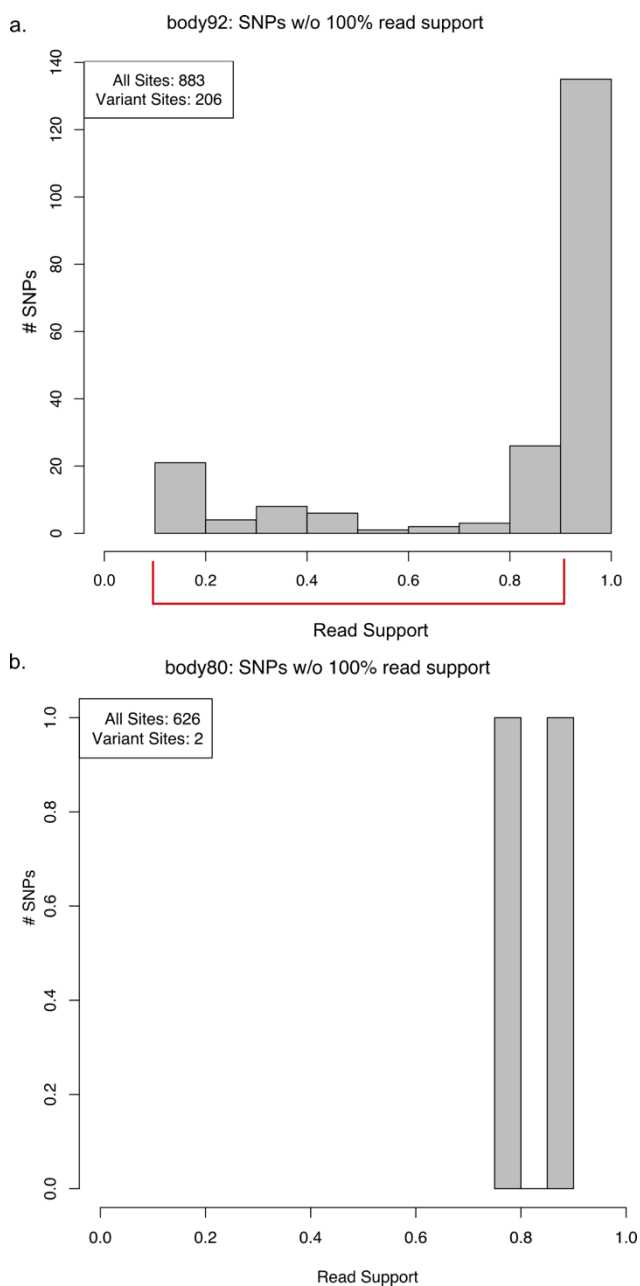


Figure S11. Heterozygosity plots for body92 and body80. A) The grey bars represent the quantity of sites at which between 10-99% of the reads represent an alternative derived allele (the allele must have coverage of 5-fold or greater to be counted). Body92 has 206 of these sites in total, though most of them have >90% representation, making them dominant alleles.

The minority alleles, falling between 10-90% of the reads representing a given site (within the red bracket), are 70 in total. B) The grey bars represent the quantity of sites at which between 10-99% of the reads represent an alternative derived allele. Body80 has only two variant positions.

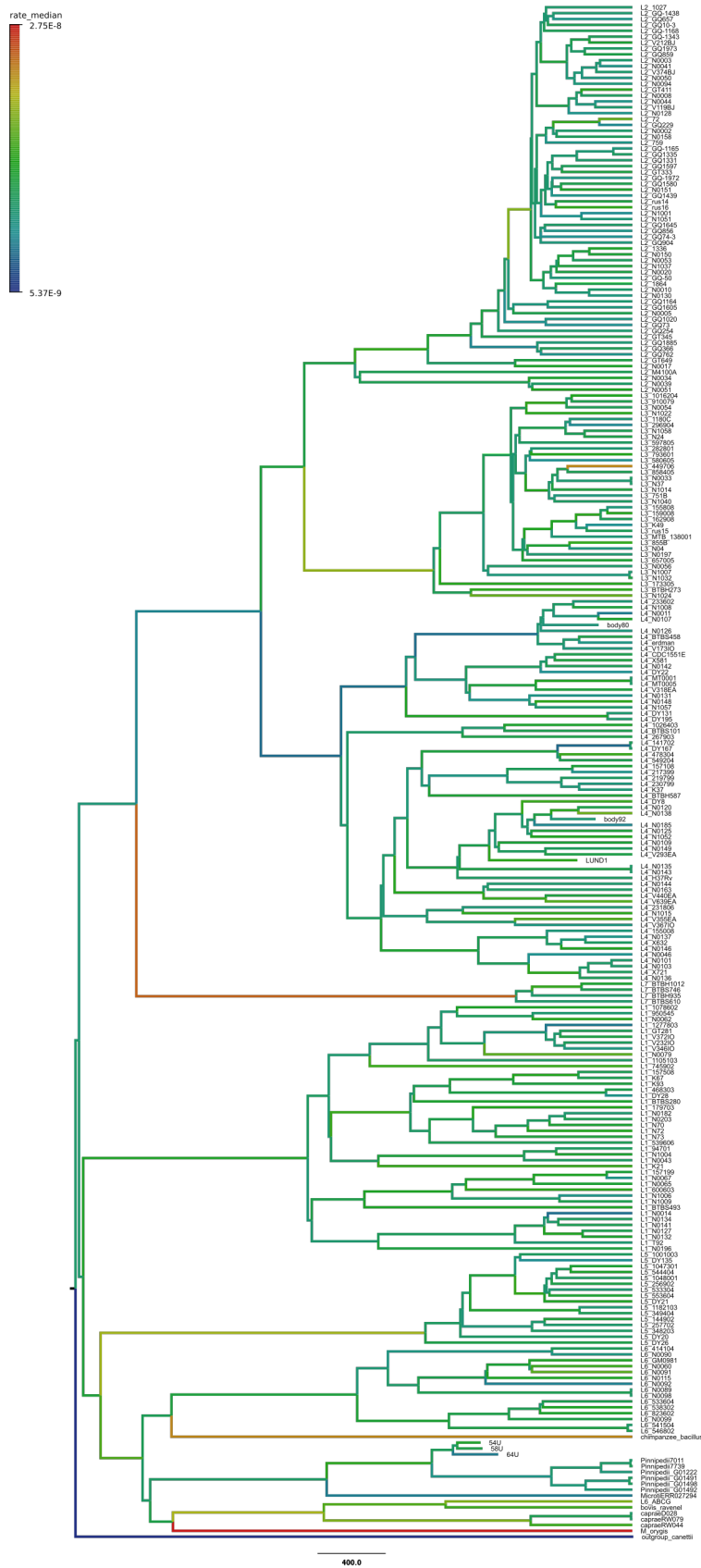


Figure S12. MTBC BDSKY+UCLD maximum clade credibility tree. Branches of the tree were colored according to median substitution rate.

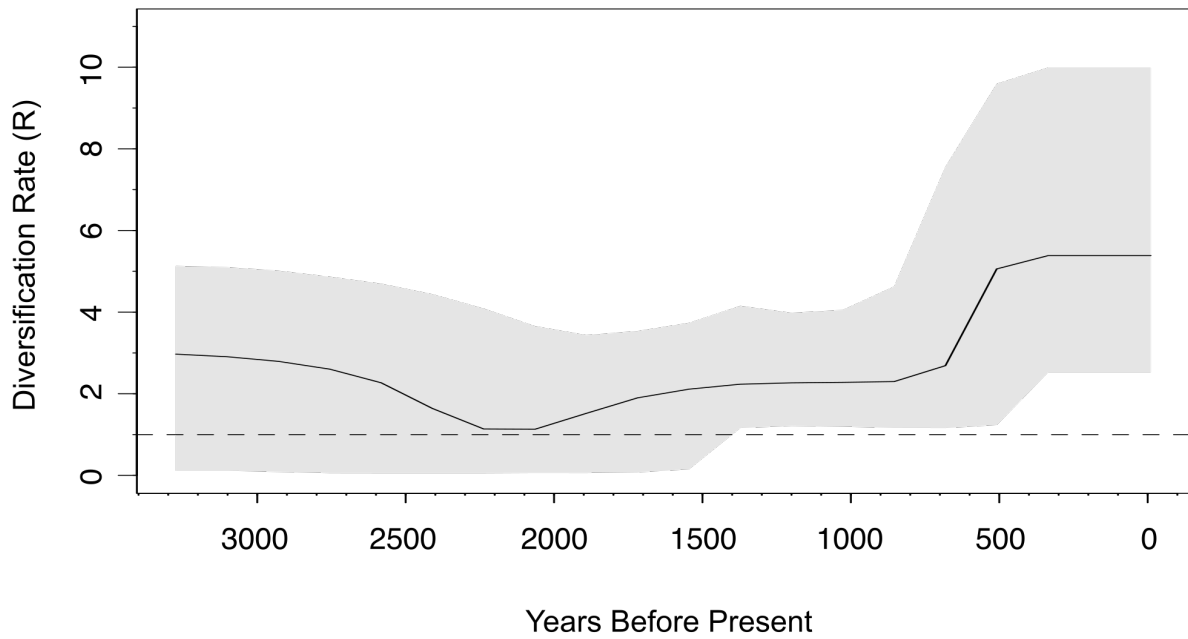


Figure S14. Full MTBC BDSKY plot. The central black line indicates mean R over time, the shaded grey area represents the 95% HPD interval of R over time, and the dashed line shows $R=1$.

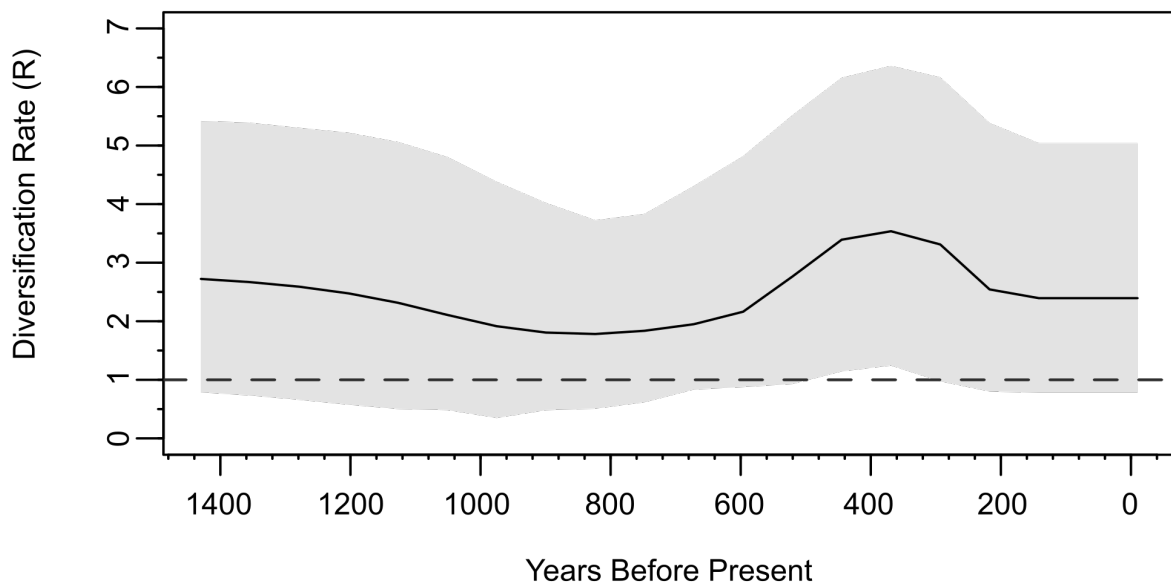


Figure S15. L4 BDSKY plot. Birth-death skyline plot from the BDSKY+UCLD+noOrigin model applied to the L4 dataset. The central black line indicates mean R over time, the shaded grey area represents the 95% HPD interval of R over time, and the dashed line shows $R=1$.

5. Manuscript B

SCIENTIFIC REPORTS

OPEN

Differential preservation of endogenous human and microbial DNA in dental calculus and dentin

Received: 20 March 2018
Accepted: 14 June 2018
Published online: 29 June 2018

Allison E. Mann^{1,2,3}, Susanna Sabin¹, Kirsten Ziesemer⁴, Åshild J. Vågane¹, Hannes Schroeder^{4,5}, Andrew T. Ozga^{1,2,3,6}, Krithivasan Sankaranarayanan^{3,7}, Courtney A. Hofman^{2,3}, James A. Fellows Yates¹, Domingo C. Salazar-García^{1,8}, Bruno Frohlich^{9,10}, Mark Aldenderfer¹¹, Menno Hoogland⁴, Christopher Read¹², George R. Milner¹³, Anne C. Stone^{6,14,15}, Cecil M. Lewis Jr.^{2,3}, Johannes Krause¹, Corinne Hofman⁴, Kirsten I. Bos¹ & Christina Warinner^{1,2,3}

Dental calculus (calcified dental plaque) is prevalent in archaeological skeletal collections and is a rich source of oral microbiome and host-derived ancient biomolecules. Recently, it has been proposed that dental calculus may provide a more robust environment for DNA preservation than other skeletal remains, but this has not been systematically tested. In this study, shotgun-sequenced data from paired dental calculus and dentin samples from 48 globally distributed individuals are compared using a metagenomic approach. Overall, we find DNA from dental calculus is consistently more abundant and less contaminated than DNA from dentin. The majority of DNA in dental calculus is microbial and originates from the oral microbiome; however, a small but consistent proportion of DNA (mean $0.08 \pm 0.08\%$, range 0.007–0.47%) derives from the host genome. Host DNA content within dentin is variable (mean $13.70 \pm 18.62\%$, range 0.003–70.14%), and for a subset of dentin samples (15.21%), oral bacteria contribute > 20% of total DNA. Human DNA in dental calculus is highly fragmented, and is consistently shorter than both microbial DNA in dental calculus and human DNA in paired dentin samples. Finally, we find that microbial DNA fragmentation patterns are associated with guanine-cytosine (GC) content, but not aspects of cellular structure.

Dental calculus is a mineralized form of dental plaque¹, a sequentially generated microbial biofilm² that entraps microbial, dietary, host, and ambient debris during spontaneous calcification events³. Unlike body mucosal surfaces that have continual cell turnover, teeth do not remodel. Consequently, they are relatively stable environments for bacterial colonization during biofilm development⁴, making the formation of dental calculus difficult to prevent without mechanical removal. As a result, dental calculus is prevalent in the archaeological record, and due to its excellent morphological preservation, it has long been an attractive target for microscopic analysis^{5–11}. More recently, dental calculus has been explored as a source of ancient DNA (aDNA) and has been shown to retain

¹Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, 07745, Germany. ²Department of Anthropology, University of Oklahoma, Norman, Oklahoma, 73019, USA. ³Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, Oklahoma, 73019, USA. ⁴Faculty of Archaeology, Leiden University, Leiden, 2333 CC, The Netherlands. ⁵Natural History Museum of Denmark, University of Copenhagen, Copenhagen, 1350, Denmark. ⁶Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, 85287, USA. ⁷Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma, 73019, USA. ⁸Grupo de Investigación en Prehistoria IT-622-13 (UPV-EHU)/IKERBASQUE-Basque Foundation for Science, Vitoria, Spain. ⁹Department of Anthropology, Dartmouth College, Hanover, New Hampshire, 03755, USA. ¹⁰Department of Anthropology, Smithsonian Institution, Washington DC, 20560, USA. ¹¹School of Social Sciences, Humanities, and Arts, University of California Merced, Merced, California, 95343, USA. ¹²Institute of Technology Sligo, Sligo, F91 YW50, Ireland. ¹³Department of Anthropology, Penn State University, University Park, Pennsylvania, 16802, USA. ¹⁴School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, 85287, USA. ¹⁵Institute of Human Origins, Arizona State University, Tempe, Arizona, 85287, USA. Allison E. Mann and Susanna Sabin contributed equally to this work. Correspondence and requests for materials should be addressed to K.I.B. (email: bos@shh.mpg.de) or C.W. (email: warinner@shh.mpg.de)

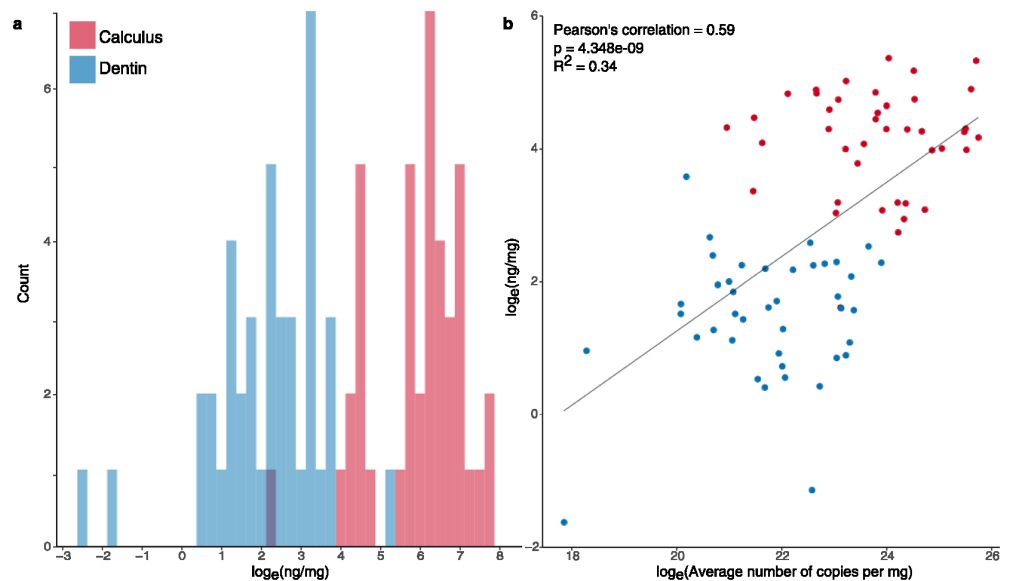


Figure 2. Total DNA content of dental calculus is higher than dentin as measured by both fluorescence and quantitative PCR (qPCR) techniques. **(a)** Normalized DNA yield (log transformed nanograms DNA per milligram starting material) of DNA extracts obtained from dental calculus and dentin as measured by a Qubit fluorometer using a High Sensitivity Assay ($p = 3.911e-11$, Wilcoxon signed-rank paired test). **(b)** Linear correlation between normalized DNA yield (log transformed nanograms DNA per milligram starting material) and normalized DNA yield as measured by qPCR after library preparation. The average number of copies per milligram was calculated from the mean of four technical replicates, and these values were compared between dentin and calculus ($p = 6.941e-07$, Wilcoxon signed-rank paired test). In both quantification metrics dental calculus has a higher DNA yield than dentin.

source of the DNA (human vs. microbial) but not with cellular structure (e.g., microbial cell wall type or presence of a surface-layer). Additionally, human DNA is consistently shorter in dental calculus than in paired dentin samples, which may reflect differences in the manner by which human DNA is incorporated into each matrix. Finally, we observe a systematic loss of short AT-rich DNA fragments that is particularly marked in bacteria with low to medium GC content genomes.

Results

DNA abundance. The total amount of DNA recovered from dental calculus is substantially higher than from dentin as measured by both fluorometric quantitation (Qubit) (Fig. 2a) and quantitative PCR (qPCR) (Fig. 2b). Immediately following extraction, DNA yields from dental calculus as measured by fluorometry ranged from 4.9 ng/mg to 214.4 ng/mg (median 72.1 ng/mg), while dentin samples from the same individuals yielded 0.2 ng/mg to 35.7 ng/mg (median 4.8 ng/mg). These measurements are moderately correlated (Pearson's coefficient: 0.59; 95% CI: 0.43, 0.72) to the average DNA copy number estimated after library construction using qPCR. This fit accounts for 34% of the total variance using a linear regression model (Fig. 2b). Differences between the two methods may be artifacts of DNA library construction, during which inefficient ligation of universal adapters to the DNA fragments and several silica-based purification steps contribute to substantial, but stochastic, DNA loss²⁷. While differences in DNA abundance are not as stark in the qPCR results, the overall amount of DNA recovered from dental calculus is significantly higher than from dentin using both metrics (fluorometry, $p = 7.822e-11$, Wilcoxon signed-rank paired test; qPCR, $p = 6.941e-07$, Wilcoxon signed-rank paired test).

Microbial community composition and contamination. Overall, archaeological dental calculus and dentin contain microbial DNA from highly distinct communities. To explore these differences, shotgun-sequenced reads from the 48 dentin and dental calculus pairs were taxonomically binned for microbial identification using MALT²⁸, and the resulting assignments were visualized using MEGAN²⁹ (Supplementary Tables S2 and S3). A species-level PCoA plot based on a Bray-Curtis dissimilarity matrix demonstrates that dental calculus samples form a relatively tight and cohesive group that is distinct from the more diffuse distribution of microbial communities identified within dentin (Fig. 3a). Alpha diversity is significantly higher for dentin samples as compared to dental calculus (Mann-Whitney-U, $p = 8.55E-15$), consistent with a signal of external contamination (Supplementary Figure S1). Differences in microbial community structure and composition between dentin, calculus, and blanks was further explored using linear discriminant analysis³⁰. Dentin and blank samples are primarily discriminated by taxa that are environmentally abundant, while calculus samples are discriminated by typically oral taxa (see Supplementary Methods; Supplementary Table S4). Importantly, microbial communities from each material (dental calculus or dentin) are less similar to their paired sample than they are to samples of the same material. This pattern is consistent with expectations that the microbial taxa within dental calculus

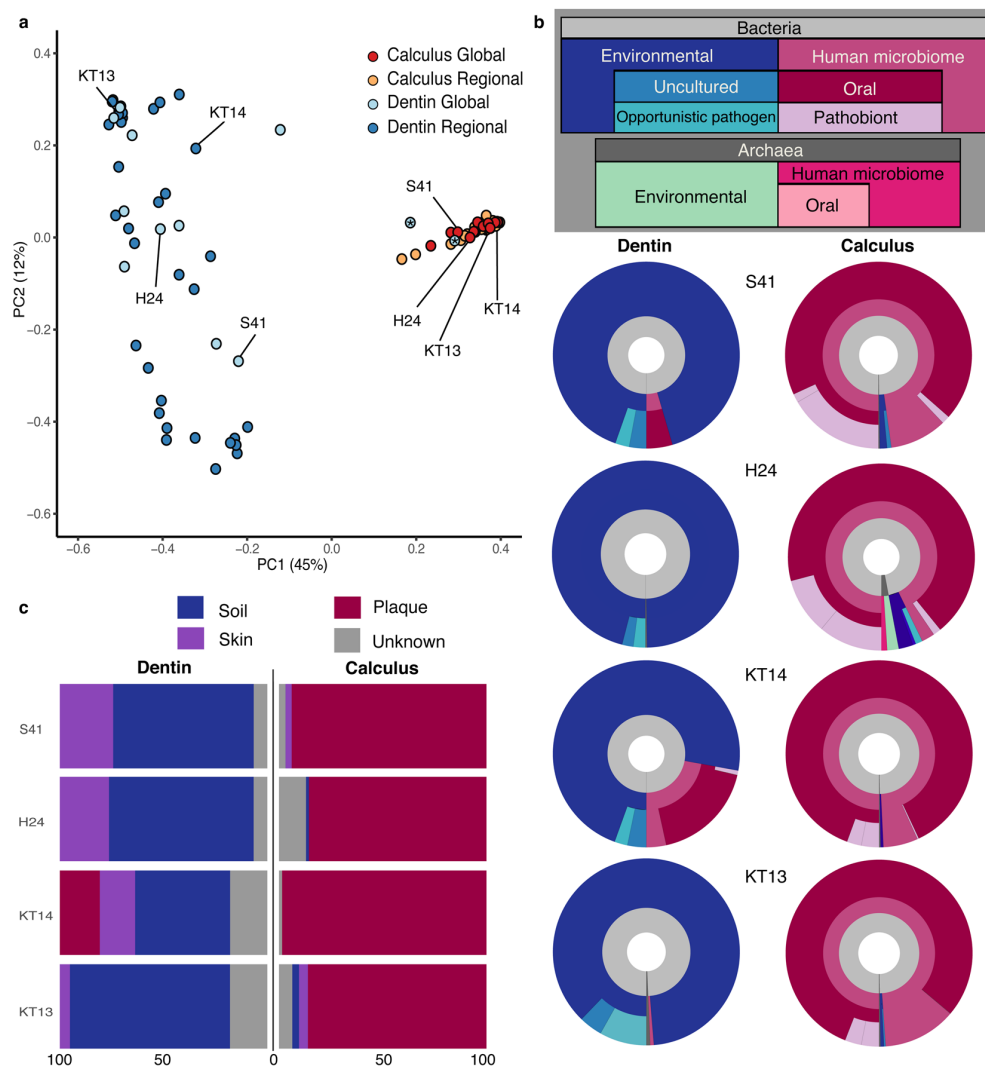


Figure 3. Microbial communities represented in archaeological dental calculus and dentin are distinct. **(a)** Principal Coordinates Analysis (PCoA) of Bray-Curtis distances of all bacterial and archaeal species-level assignments from dental calculus and dentin. Color indicates material type (dental calculus or dentin) and membership in global or regional dataset. Microbial taxonomy was assigned using MALT with the full NCBI nucleotide database. Dentin samples marked with an asterisk belong to individuals NF47 and NF217 in the global dataset, and may represent teeth with carious lesions (Supplementary Figure S2). **(b)** A subset of four sample pairs from the global (S41, H24) and regional (KT14, KT13) sample sets were selected to further demonstrate differences in dental calculus and dentin microbial communities. Species represented in the MALT results were sorted into a layered classification scheme (see legend in gray box), and the proportions of reads assigned to each taxon were used to generate Krona plots. **(c)** Stacked bar plots of Bayesian SourceTracker results for the four selected pairs display estimated proportions of source contribution at the genus level, using modern plaque, skin, and soil datasets as model sources. Both approaches show overwhelming abundance of environmental bacteria within dentin samples, while most microbial DNA within the paired calculus samples are associated with, and most likely derive from, the human oral microbiome. The robustness of the human microbiome signal to the environmental signal is driving the separation we see in the PCoA.

represent a relatively well preserved biological community derived from dental plaque, while dentin – being sterile in life – is expected to harbor a microbial community primarily composed of exogenous contaminant bacteria acquired through stochastic postmortem processes.

Two additional approaches were used to characterize further the nature and inferred sources of the microbial communities present within sample-types. First, microbial species-level identifications were categorized according to a nested scheme reflecting organism membership in one or more of the following source categories: environmental, uncultured environmental, human microbiome, human oral, pathobiont, and opportunistic pathogen (Fig. 3b; Supplementary Table S5), which were then visualized using the Krona Excel template³¹. Category

membership was determined by species presence or absence in the Human Oral Microbiome Database³² as well as source and habitat descriptions in the twenty most recent articles in PubMed³³ using the species name as the search keyword. Using this analysis, stark differences are observed in the inferred microbial source contributions to archaeological dental calculus and dentin, whereby dental calculus is strongly dominated by human-associated – and especially oral-associated – taxa, while dentin is primarily composed of environmental taxa (Supplementary Table S6).

As a separate confirmation method, SourceTracker, a Bayesian source-prediction tool³⁴, was used to estimate proportions of source similarity (Supplementary Table S7) in dental calculus and dentin based on a set of modern reference microbial communities sequenced from human dental plaque^{35,36}, human hand swabs³⁷, and top soil samples from Oklahoma and Alaska³⁸ (Fig. 3c). The human hand swab reference community was included in this analysis to detect microbial contaminants introduced through handling during excavation or storage. In agreement with the other methods presented here, archaeological dental calculus is estimated to be composed primarily of dental plaque-associated taxa, while dentin is dominated by genera associated with human skin and environmental sources (Supplementary Table S8). The highest predicted contributions of skin and soil to dental calculus samples are 7.9% and 14.0%, respectively, while the same predicted contributions for dentin are 33.5% and 89.0%, respectively. Together, these analyses suggest dental calculus is relatively robust to environmental contamination when compared to dentin.

Although most dentin samples are strongly dominated by environmental taxa, two dentin samples from the site of Norris Farms – NF47 and NF217 – cluster with dental calculus samples in the PCoA (Fig. 3a) and are estimated via SourceTracker analysis³⁴ to contain microbial DNA that is 56.2% and 78.4% derived from dental plaque, respectively (Supplementary Figure S2). Unlike other teeth in this study, the Norris Farms teeth were obtained in a fragmented state. Lacking the full intact teeth, the presence of carious lesions could not be ruled out. In addition, because caries are not associated with the presence, absence, or even abundance of specific organisms, we were not able to determine whether a carious lesion was present using metagenomic data alone^{39,40}. For this reason, the Norris Farms samples were excluded from further downstream analyses.

In samples for which the tooth was intact, tooth dentin is generally strongly dominated by environmental taxa; however, a subset of dentin samples exhibit a slight signal of the human oral microbiome, ranging from 0.0% to 40.0%, with 7 of 46 dentin samples having a predicted oral source contribution of 20.0% or more by SourceTracker analysis (Supplementary Figure S3). A PCoA and SourceTracker analysis including blanks confirms this oral signal is not due to laboratory-based contamination (Supplementary Figure S4). Based on our classification scheme visualized by Krona plots, microbial DNA derived from human microbiome sources constituted a median of 98% (ranging from 90.1% to 99.3%) of the total microbial contribution in the calculus samples, whereas, for dentin, human microbiome DNA constitutes a median of 3.7% (ranging from 0.2% to 61%; Supplementary Table S6).

Human DNA content. Although typically higher than in dental calculus, the proportion of human endogenous DNA in dentin varies substantially, ranging more than 3 orders of magnitude in this study, from 0.034% to 70.1% of all reads (Fig. 4a). By contrast, the proportion of human DNA in dental calculus is relatively low, but consistent across all samples, differing by less than 2 orders of magnitude, from 0.007% to 0.47%. To verify these reads as authentic host DNA and to mitigate the possibility that they represent spurious mapping to the human genome or modern contamination, a secondary verification procedure was performed. Here, only those reads that (1) met stringent mapping criteria when mapped to the hg19 human reference genome (see Supplementary Methods), (2) were assigned to the *Homo sapiens* node in lowest common ancestor assignment by MALT, and (3) displayed typical ancient DNA damage profiles⁴¹ were included in a high confidence human dataset for further inspection.

The number of human-assigned reads following strict mapping decreased across all samples but was more severe among samples from the Kiltasheen (regional) dataset. As the strict mapping parameters allow only one mismatch per 50 base pairs, this comparatively high loss of reads in the regional dataset likely results from the fact that these samples were sequenced using a single-end, 75 cycle sequencing strategy, rather than the paired-end 200 cycle sequencing strategy employed for the global sample set. Of all originally designated human reads in the regional dataset, between 20.7% to 60.9% of the dental calculus reads and 0.6% to 67.1% of dentin reads pass the initial strict mapping step. Within the global dataset, between 60.1% to 88.3% of dental calculus and 74.0% to 97.9% of dentin reads pass.

Reads passing strict mapping were next analyzed with MALT using the full NCBI nucleotide database as a reference to ensure proper assignment to *Homo sapiens* rather than to other organisms. Among all dental calculus samples, the proportion of reads uniquely assigned to the *Homo sapiens* node ranges from 79.6% to 90.3%. Within all dentin samples the proportion of reads assigned to *Homo sapiens* ranges from 81.1% to 92.7%.

Finally, to establish that these reads are consistent with patterns expected of authentic ancient host DNA and not modern contamination, rates of terminal cytosine deamination were evaluated using mapDamage 2.0⁴¹. While damage rates declined (median 3% in dentin and 4% in calculus) after these verification steps (Wilcoxon sign rank test, dentin $p = 1.36E-15$, calculus $p = 6.74E-15$; See Supplementary Table S9), all samples except one dentin (KT05, Supplementary Figure S5d) present damage patterns consistent with authentic aDNA both pre- and post-verification. This contrasts with contaminant human DNA observed in the extraction and library blanks, which showed no evidence of DNA damage, and thus is consistent with modern human contamination (Supplementary Figure S5; Supplementary Table S10). Because the human DNA sequences present in the blanks lack damage, they are therefore unlikely to be the source of the damaged human reads in the dentin and calculus samples. Taken together, these analyses suggest that although a small subset of reads may be erroneously assigned to the human genome, most reads are likely correctly assigned, and these reads appear to be authentic ancient human DNA and not contamination.

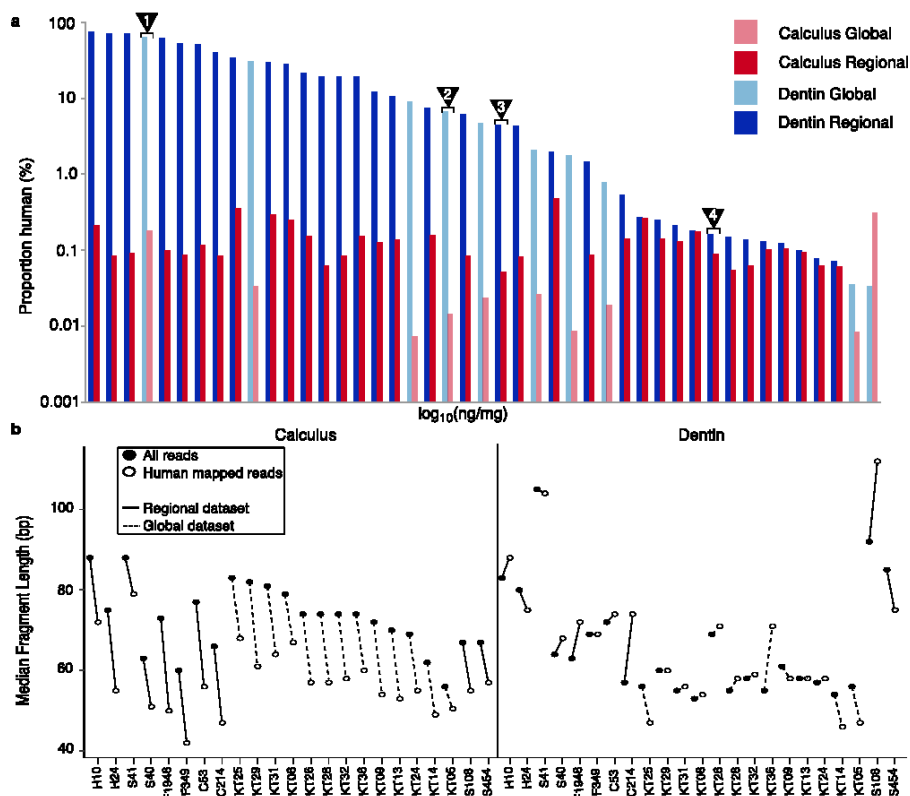


Figure 4. Human DNA in dental calculus shows consistent patterns of low relative abundance and high fragmentation. (a) Relative percentage of human DNA in all paired calculus and dentin samples calculated from de-duplicated reads mapped to the hg19 human reference genome using BWA. While the majority of dentin samples have an overall higher percentage of human DNA, this value varies substantially by sample. Calculus is comparatively consistent between samples albeit on average lower than their paired dentin sample. Sample pairs corresponding to those in Fig. 3 are indicated by numbered triangles: (1) S41, (2) H24, (3) KT14, (4) KT13. (b) Median fragment length of merged reads mapping to the human genome compared to all merged reads in both dental calculus and dentin. Human assigned reads in dental calculus are shorter than expected independent of age, laboratory processing protocol, or sample preservation. Human mapped reads in dental calculus and dentin were further verified for authenticity using strict mapping parameters (Supplementary Figure S5, Supplementary Table S9).

Next, DNA fragmentation and damage patterns of human reads in paired dentin and dental calculus samples were compared for a subset of the samples for which paired-end DNA sequence data was generated (Supplementary Table S11), which includes the entire global sample set ($n = 10$ individuals) and a representative subset of the regional sample set ($n = 13$ individuals). Overall, the median fragment lengths of total DNA recovered from both dental calculus (56–88 bp, mean 72.8 bp) and dentin (53–105 bp, mean 66.0 bp) are short and fall within a size range expected for archaeological samples (Supplementary Table S1). For each dental calculus sample, the median fragment length of human reads was found to be 15.5 ± 4.2 bp shorter than the overall median fragment length of DNA in each sample, which is primarily microbial in origin (Fig. 4b). Dentin samples, however, show no pattern with respect to human DNA fragment length compared to overall DNA fragment length. Because some dentin samples have a high endogenous human content, a separate analysis comparing the fragment length profile of human mapped reads and all non-human reads was performed (Supplementary Figure S6). In calculus, there is a greater average magnitude of difference between the mean fragment length of human and non-human reads (calculus: 0.50 ± 0.15 ; dentin: 0.20 ± 0.17 , Cohen's d), and the median length of non-human reads is consistently longer than that of human reads. Conversely, the magnitude of difference between human and non-human fragment length profiles in dentin is variable and inconsistent (Supplementary Table S11). This analysis confirms that the observed differences in fragment length profiles between dentin and dental calculus is not the result of high human endogenous content of particular samples. Comparing human DNA in dental calculus and dentine, we find that human DNA within dental calculus is generally more fragmented than human DNA in paired dentin samples, with the median length of calculus-derived fragments being approximately $10.3 \text{ bp} \pm 12.0$ shorter than that of dentin-derived fragments (Wilcoxon signed-rank test, $p = 0.01178$); however, this pattern is largely driven by the relatively long human DNA fragment lengths in dentin, and further work is needed to determine if this is an artifact of sample preparation or a true biological pattern.

We next investigated the relative degree of terminal cytosine deamination among human reads in dental calculus and dentin pairs from the same individual (Supplementary Figure S7). Across all samples, the terminal cytosine deamination rate of dentin is significantly higher than dental calculus ($p = 0.0009$, Wilcoxon signed-rank test), yet there is no significant difference in the cytosine deamination in overhangs (δ_o , $p = 0.49$, Wilcoxon signed-rank test) or in double stranded regions (δ_s , $p = 0.07$, Wilcoxon signed-rank test). However, the average length of overhangs (λ) is significantly different between all dentin and calculus samples ($p = 0.04$, Wilcoxon signed-rank test) (Supplementary Table S9). These results suggest that while differential damage parameters may not be consistent between paired samples, some patterns may be predicted by sample type.

Microbial DNA fragmentation and damage patterns. We next investigated fragmentation and terminal cytosine damage patterns for a selection of oral microbes preserved at high abundance within dental calculus in order to determine the impact of cell wall or genomic structure on aDNA preservation (Supplementary Table S12). It has been previously suggested that cell wall composition may influence the preservation of microbial DNA in archaeological dental calculus and dentin^{13,42}. Fifty oral bacteria were selected based on their frequency in the dataset and presence in the Human Oral Microbiome Database³² and categorized into groups based on whether they are known to be Gram stain positive or Gram stain negative, the presence or absence of a surface layer (S-layer), and overall genomic GC content (Figs 5a and c). Additionally, a subset of 20 of these taxa that were highly abundant in our dataset was analyzed to examine species-level patterns of fragmentation (Fig. 5b) and DNA damage (Fig. 5d).

We found no indication of a relationship between terminal cytosine damage and microbial genomic source (Fig. 5c,d, Supplementary Table S13) nor a relationship between fragmentation and cell wall structure. However, we do see a small decline in average DNA fragment length in taxa with higher genomic GC content (Fig. 5a). This pattern is also reflected among the 20-species subset of oral bacteria chosen for closer analysis, and reads assigned to *Actinomyces radidentis*, the species with the highest GC content, have the largest displacement from the median fragment length of all selected taxa (Fig. 5b).

GC content shifts. Finally, the relationship between fragment length and mean GC content was examined for five prevalent oral genera in dental calculus samples and a single prevalent soil genus in dentin samples to evaluate further the influence of genomic structure on microbial DNA survival. Genera were chosen to maximize the range of GC content with two genera each representing low, medium, and high-expected genomic GC content (Fig. 6). Among all published genomes available in the NCBI genome database³³, members of the genus *Methanobrevibacter* range in genomic GC content from 24.2% in *M. wolinii* to 32.6% in *M. ruminantium*. The common oral methanogen *M. oralis* is expected to have a GC content of 27.9%. Members of *Fusobacterium* also have low GC content, ranging from 26.0% (*F. perfoetens*) to 35.1% (*F. necrophorum*). Among the medium GC content genera, the GC content of *Tannerella* ranges from 37.7% (*T. sp.* CAG:118) to 56.5% (*T. sp.* oral taxon HOT-286), and *Porphyromonas* range from 42.7% (*P. gingivicanis*) to 56.3% (*P. benmonis*). Finally, for high GC content genera, *Actinomyces* ranges from 49.6% (*A. coleocanis*) to 73.1% (*A. dentalis*), and *Streptomyces* ranges from 56.4% (*S. sp.* WAC00263) to 71.1% (*S. sp.* NBRC110027). In comparing reads assigned to these genera in our samples, we detect an increase in GC content at shorter read lengths among all chosen genera except for *Streptomyces*, where no shift was observed. Importantly, this shift is greater for genera with moderate and low expected GC content. For example, in *Methanobrevibacter* and *Fusobacterium* the length at which the mean GC content begins to substantially shift (≥ 1 z score) is 39 bp and 48 bp, respectively (Supplementary Figure S8). This shift occurs at longer lengths for *Tannerella* and *Porphyromonas* at 59 bp and 58 bp, respectively, while *Actinomyces* does not present a substantial shift until 35 bp. No shift is observed in *Streptomyces*, although this may occur in short fragments that are below the length-filtering threshold (30 bp) used for this dataset.

Discussion

Dental calculus is a richer source of genetic material than dentin. In agreement with the findings of previous studies^{14,17}, overall DNA recovery from ancient dental calculus was found to be substantially higher than from dentin, and this pattern is consistent through time and across preservation contexts. This higher DNA content of archaeological dental calculus compared to dentin likely reflects biological differences between the two substrates in cellular composition and structure during life, as well as decomposition patterns after death.

Dental calculus is formed from dental plaque, a dense microbial biofilm that has been estimated to contain more than 200 million cells per milligram⁴³. Approximately 70% of the dry weight of plaque consists of microbial cells⁴⁴, and a large proportion of the biofilm matrix itself is composed of extracellular bacterial DNA, which provides both structural support and protection to its microbial inhabitants⁴⁵. Furthermore, the mineralization process that leads to dental calculus formation involves rapid inter- and intracellular crystal formation by calcium phosphates, including hydroxyapatite, which strongly bind DNA. The result is a dense crystalline structure that is relatively inert and resistant to microbial attack, enzymatic action, and non-acidic chemical alteration⁴⁶. Although the surface of dental calculus remains porous⁴⁷, penetration of substances towards the internal layers of the calculus matrix is restricted⁴⁸, which may account for its high DNA preservation qualities. Calculus is not an entirely closed system, however. It has been shown to disproportionately lose soluble small metabolites over time⁴⁹, suggesting that the mineralized matrix allows some degree of water movement.

In contrast to dental calculus, dental hard tissues are largely acellular, with live cells in mature teeth being limited to a layer of odontoblasts lining the pulp chamber wall, a sparse distribution of entrapped cementocytes within apical cementum, and a layer of cementoblasts around the periodontal ligament^{44,50,51}. Most cells within teeth are instead found within the dental pulp and consist of perivascular cells, blood cells, and pulpal blood vessels^{44,52}, all of which decompose readily after death through a combination of necrosis and microbial invasion³. Thus, while the majority of cells within dental calculus are found within a mineralized structure conducive to

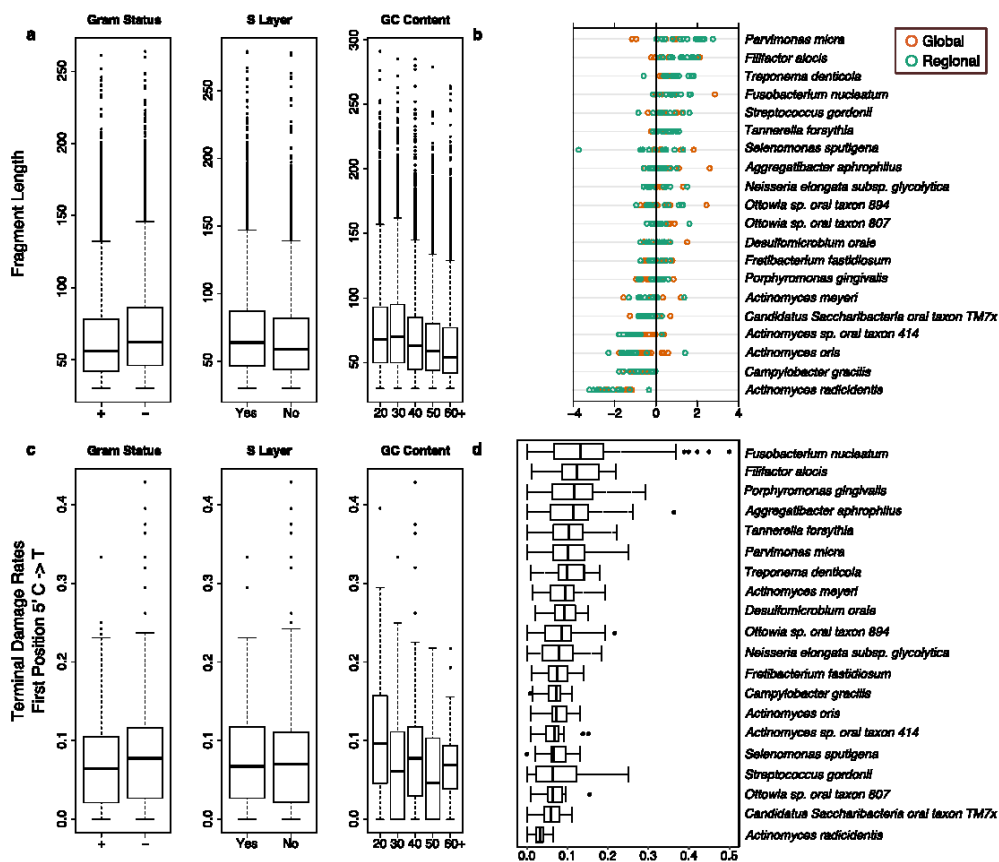


Figure 5. Fragment length and damage rates among bacterial taxa within calculus. (a) Fragment length distribution of 50 high frequency species-level bacteria among all dental calculus grouped into three metadata categories: gram status, the presence or absence of a surface layer (S-layer), and the overall genomic GC content of the organism as documented from the reference genome in the NCBI database. Input was normalized to 400 randomly chosen reads per sample per species to mitigate the impact of sample specific read length profiles. No detectable differences in the median are apparent for the presence of an S-layer or gram status. Genomic GC content, however, does impact the overall fragment length profile of reads. (b) Deviation of the median fragment length from overall sample median fragment length of a subset of 20 oral bacteria in dental calculus colored by dataset origin. Again, genomic GC content appears to impact the fragment length profiles of individual taxa where high GC content species have shorter overall fragment length profiles. (c) Terminal cytosine damage rates (C to T substitution ratio at the first position of the 5' end of the molecule) among 50 bacterial species in dental calculus grouped by gram status, presence or absence of an S-layer, and overall genomic GC content. No detectable differences in damage rates could be detected in any of the three chosen metadata categories. (d) Terminal cytosine damage rates (C to T substitution ratio at the first position of the 5' end of the molecule) among 20 oral bacteria. While some differences in damage patterns can be detected, it is unclear if this is random or due to some process not examined in this study.

preservation, the majority of cells within teeth are not, which may partially explain the large differences in total DNA yield between the two substrates. However, further studies of total DNA yields from freshly extracted teeth and their component tissues are needed to fully understand these differences.

Dental calculus and dentin harbor distinct microbial communities. Microbial DNA obtained from dental calculus and dentin derives from distinct communities. In agreement with previous studies^{3,14}, the microbial community in dental calculus is dominated by human-associated oral taxa, and DNA derived from these organisms greatly exceeds that originating from environmental sources. The consistent preservation of a strong oral microbiome signal in all 48 dental calculus samples in this study suggests that this pattern is typical for archaeological dental calculus. By contrast, microbial DNA within dentin primarily derives from environmental sources. This distinction between the two substrates is preserved across geography, burial environment, and temporal period.

The tight clustering of all calculus samples included in this study, in contrast to the diffuse distribution of dentin samples in the PCoA (Fig. 3a), indicates that the oral microbiome signal is relatively uniform and stable across diverse contexts, as expected for a preserved biological community. In comparison, the diffuse distribution

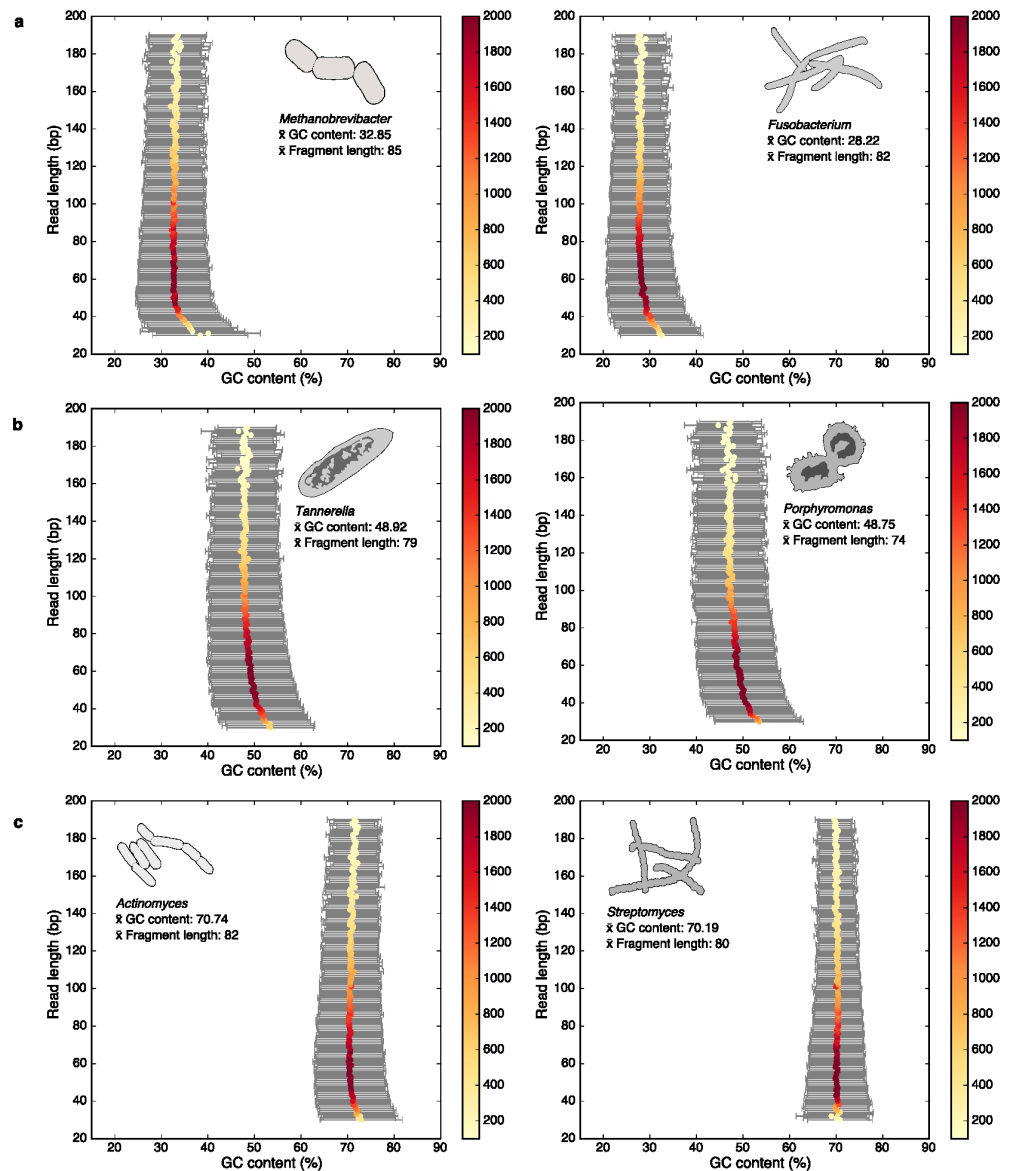


Figure 6. Relationship of GC content to fragment length in five prevalent oral genera and one soil genus (*Streptomyces*). (a) Two low expected GC content genera, *Methanobrevibacter* and *Fusobacterium* binned by read lengths wherein each dot represents the mean GC content for that bin. (b) Two moderate expected GC content genera, *Tannerella* and *Porphyromonas* binned by read lengths wherein each dot represents the mean GC content for that bin. (c) One high expected GC content oral genus (*Actinomyces*) and one high expected GC content soil genus (*Streptomyces*) binned by read lengths wherein each dot represents the mean GC content for that bin. Heat bar indicated the number of reads representing each length bin. In all cases except for *Streptomyces* only reads from dental calculus samples were included. Reads for *Streptomyces* were extracted from dentin samples to contrast GC shifts in a potentially modern soil contaminate with typical members of the human oral microbiome. In all cases except for *Streptomyces* the GC content of shorter reads is skewed higher than the overall mean value. This effect greater in low to moderate expected GC content genera than those with higher GC content (Supplementary Figure S8).

of dentin samples reflects the diverging influences of exogenous microbes from different environmental contexts and the absence of a consistent microbial composition. Despite the presence of DNA belonging to oral microbes in some dentin samples, none of those included in this analysis join the calculus cluster, indicating they are the result of stochastic preservation of particular oral microbes and do not retain a signal of a biological community.

Dentin is a source of oral microbial DNA. Although most microbial DNA within dentin is environmental in origin, we find that oral bacteria contribute >20% of total DNA in 7 of 46 dentin samples in this study. Notably, the teeth in this study (with the exception of the two excluded Norris Farms teeth) could be confirmed to be free of oral pathology, such as caries, and the dentin collection procedure included either surface cleaning or only the pulp chamber was sampled. Thus, dental infection and incomplete calculus removal are unlikely explanations for the presence of oral microbial DNA we observe in dentin. Our results agree with recently reported data⁵³, in which human-associated microbes constituted ~15% of the organisms identified in a metagenomic study of over 100 archaeological, root-derived dentin samples. These findings suggest that members of the oral microbiome may also participate in postmortem dental decomposition, although to a lesser extent than environmental microbes.

The presence of DNA from oral taxa in dentin has important implications for the study of ancient commensal microbes and their evolution. Although prevalent, dental calculus is not always present or preserved in archaeological skeletal collections. Additionally, dental calculus may be absent or found in low abundance in young individuals or for certain populations or time periods due to genetic, dietary, or behavioral differences. If insufficient calculus is available for study, it may be possible to instead access aspects of the oral microbiome through tooth dentin. Although the stochastic processes of postmortem microbial growth would preclude oral microbiome community-level analyses, genetic sequencing of dentin could nevertheless provide access to the genomes of individual oral taxa for analysis.

Dental calculus is a source of host DNA. Although dentin generally contains a higher proportion of total human DNA than dental calculus, many dental calculus samples in this study have comparable proportions of human DNA to their dentin pair, with one dental calculus sample from the Netherlands (S454) having a higher proportion of DNA assigned to the human genome than its paired dentin sample (Supplementary Table S9). This pattern is largely driven by the high variability of human DNA preservation within dentin.

Except in cases of blood-borne infections that infiltrate the pulp chamber and expose pathogen DNA to the dentin matrix⁴², nearly all DNA in dentin should originate from the host genome at the time of death. However, archaeological teeth typically contain low proportions of human DNA due to postmortem degradation and exogenous microbial growth. By contrast, the relative proportion of human DNA is uniformly low and relatively consistent in dental calculus. When dentin is strongly degraded and the relative proportion of host DNA in dentin is very low (<0.1%), the absolute amount of human DNA within dental calculus can exceed that of dentin. In such cases, the genetic richness of dental calculus appears to compensate for its low relative proportion of human DNA. Although obtaining host DNA from dental calculus using shotgun sequencing is generally inefficient given its low relative abundance, dental calculus has been shown to be a valuable reservoir for recovering host DNA using DNA capture methods¹⁷.

Human DNA in dental calculus is highly fragmented. We find human DNA from dental calculus to be consistently shorter than the total DNA from the same sample, and on average shorter than human DNA recovered from the paired dentin sample. As a microbial biofilm, dental calculus is not a human tissue and does not contain viable human cells. The mechanisms by which human DNA is incorporated into dental calculus are not well understood but are presumed to include passive adsorption of human DNA from oral fluids and shed mucosal cells, as well as more active incorporation through host inflammatory responses, including a kind of immune response mediated by neutrophils known as NETosis^{14,54}. Neutrophils, the most abundant nucleated cell type in human blood, are an essential cell type of the innate immune system that are recruited during active microbial infection⁵⁵. Particularly important in the pathogenesis of periodontal disease, neutrophils are recruited in high numbers into the gingival crevice to attack dental plaque bacteria^{45,56}. Previous research has found that most human proteins recovered from both modern and ancient dental calculus are associated with the innate immune system, and specifically with neutrophils¹⁴.

If host immune activity is a major contributor of human DNA to dental calculus, the role of neutrophils in this activity may partially explain the higher degree of human DNA fragmentation in dental calculus than in dentin. While neutrophils and other immune cells are capable of phagocytizing individual or small aggregates of microbial cells, large pathogens or those that can thwart phagocytosis by forming biofilms stimulate the formation of neutrophilic extracellular traps (NETs)^{57,58}. NETs are composed of decondensed chromatin that is released from the nuclear membrane and mixed with disarticulated histones and granules containing antimicrobial proteins before being ejected from the cell^{55,58–60}. The expelled chromatin traps the offending microbes while simultaneously promoting destruction of the entrapped cells⁵⁵. Interestingly, along with the granular proteins, disarticulated histones and short fragments of DNA (<100 bp) are also potent antimicrobials, likely increasing the antibiotic effect of NETs^{59,61}.

Many bacteria have been shown *in vitro* to stimulate NETosis, including the oral bacteria *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans*⁶². Moreover, certain bacteria subvert NETosis by producing extracellular nucleases (DNases) which are either bound to the cell membrane or secreted from the cell⁶³. These nucleases free trapped bacteria by degrading the DNA backbone of the excreted NETs⁵⁵. This activity is particularly prevalent during periodontal disease, and a wide range of oral bacteria including *P. gingivalis*, *Tannerella forsythia*, *Fusobacterium nucleatum*, and *Parvimonas micra* are able to produce extracellular nucleases⁴⁵.

If host DNA is incorporated into dental calculus in an acellular form – either through NETosis or by another mechanism – the exposed DNA would be vulnerable to a variety of damaging processes, including both hydrolysis and extracellular nuclease activity. This may explain why, even within the same sample, human DNA within dental calculus exhibits a higher level of fragmentation than DNA derived from oral microbes⁶⁴.

Cell wall structure is not correlated with microbial DNA fragmentation or damage. It has been previously proposed that certain microbial cell wall attributes, such as the presence (Gram-positive) or absence (Gram-negative) of a thick peptidoglycan layer, may influence the preservation of microbial DNA and therefore contribute to biases in taxonomic analyses of archaeological dental calculus¹³. However, a subsequent investigation of four dental calculus samples failed to find such a correlation²⁶. In this study, we test this hypothesis in 48 dental calculus samples and find no relationship between attributes such as cell wall peptidoglycan structure or the presence of an S-layer and DNA fragmentation or terminal cytosine damage patterns. This analysis does not preclude other aspects of cellular structure (e.g., spore formation or the presence of mycolic acids) that may impact aDNA preservation and were not tested in this study. Our analysis of reads assigned to specific bacteria suggests that fragmentation and damage patterns may be taxonomically structured. However, the consistency and biological basis of these patterns is beyond the scope of the data presented here.

Loss of short AT-rich DNA fragments may contribute to taxonomic skew. Analysis of the relationship between genomic GC content, DNA fragment GC content, and DNA fragment length reveals an inverse relationship between DNA fragment length and GC content in taxa with low- and medium-GC genomes, suggesting a systematic loss of short AT-rich fragments. Short DNA fragments lack thermostability and are easily lost through denaturation. The melting temperature of short DNA fragments is primarily dependent on DNA sequence and length, in addition to environmental conditions and other factors⁶⁵, and in general sequences with longer lengths and higher GC content have higher melting temperatures. Short DNA fragments from taxa with lower GC content genomes are expected to be more susceptible to loss through denaturation because their melting point for a given fragment length will be lower, and this may contribute to taxonomic skew. Consistent with this hypothesis, we found that high GC-content genera had slightly shorter median fragment lengths overall (Fig. 5b), which accords with a higher retention of short DNA fragments.

Although we observe these patterns in archaeological dental calculus, it is unclear if this is an artifact introduced during sequencing preparation or a naturally occurring taphonomic process. Importantly, this effect is weaker or absent from high genomic GC content genera (Fig. 6), which include many soil bacteria⁶⁶. If the loss of short AT-rich fragments is primarily taphonomic and not methodological in nature, greater taxonomic skew may be expected in libraries generated using a single-stranded library preparation, which is known to retain a higher proportion of shorter DNA fragments than the double-stranded DNA library preparation method used in this study⁶⁷. Documentation of potential taxonomic-specific biases in recovery of DNA is critical as they impact downstream interpretations of metagenomic data, affecting accurate description of these ancient microbial ecosystems.

In this study, we use metagenomic sequencing data to explore patterns of preservation in microbial- and host-derived DNA in a large and diverse set of paired archaeological dentin and dental calculus samples ($n = 48$ individuals, $n = 96$ samples). Our results confirm that dental calculus is a rich source of well-preserved oral microbiome DNA and a consistent source of highly fragmented and low abundance human DNA. We find that cell wall structure has no significant association with microbial DNA preservation, but that all samples exhibit systematic loss of short AT-rich DNA fragments, a trend that disproportionately affects taxa with low and moderate GC content genomes. Finally, we show that approximately one third of teeth retain DNA from the oral microbiome and thus dentin may serve as an alternative source of oral bacterial DNA in the absence of preserved dental calculus deposits. Our findings demonstrate that dental calculus provides a favorable environment for long-term DNA preservation.

Methods

Samples and DNA extraction. Paired dental calculus and dentin samples were obtained from seven geographically and temporally distinct sites: Middenbeemster in the Netherlands ($n = 2$, ca. 1611–1866 CE)⁶⁸, Camino del Molino in southeastern Iberia ($n = 2$, cal. 2340–2920 BCE)⁶⁹, Samdzong in Nepal ($n = 2$, cal. 400–650 CE)^{70,71}, Arbulag Soum, Khövsgöl in Mongolia ($n = 2$, 930–1650 BCE)⁷², Anse à La Gourde in Guadeloupe ($n = 2$, cal. 975–1395 CE)²⁶, Norris Farms in IL, USA ($n = 2$, ca. 1300 CE)^{17,73}, and Kiltasheen in Ireland ($n = 36$, ca. 600–1300 CE)⁷⁴ (Supplementary Table S1). The first six sites were selected to represent global patterns of DNA preservation across diverse environments, burial contexts, and time periods. Remains and data from this site are referred to as the global dataset. A more extensive sampling of a single site in Kiltasheen, Ireland was performed to account for regional DNA preservation between individuals with similar burial contexts across a time transect of approximately six centuries. Remains and data from this site are referred to as the regional dataset.

Global sample set. Samples were prepared for sequencing in a dedicated ancient DNA laboratory at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR) in Norman, Oklahoma, USA. With the exception of the previously processed Norris Farms samples (see Ozga *et al.*¹⁷), teeth were first decontaminated with bleach, then the calculus was separated using a dental scaler. The crown was separated from the root using a Dremel rotary tool. Partitioned tooth roots and calculus were further decontaminated via exposure to UV irradiation for 2 minutes. DNA extraction was performed as described by Ziesemer *et al.*²⁶. Approximately 10–20 mg of dental calculus and 100 mg of dentin were crushed and then immersed in 1 ml of 0.5 EDTA for 15 minutes to remove any additional surface contaminants. Dental calculus samples were demineralized in a solution of 0.45 M EDTA and 10% proteinase K (Qiagen, the Netherlands) at 37–55 °C for 8–12 hours followed by room temperature incubation for five days until digestion was complete. Dentin samples were demineralized at room temperature. After 2 days, the EDTA supernatant was removed and refreshed with new EDTA and 50 μ l of proteinase K (Qiagen, the Netherlands). Dentin samples were then left to demineralize for an additional 3 days at room temperature. The two dentin fractions were combined after digestion was complete. Prior to demineralization, all samples were centrifuged and the supernatant was used for DNA extraction using a

phenol-chloroform-isoamyl alcohol (25:24:1) along with three blanks. Extracted DNA was isolated using a modified Qiagen MinElute silica spin column purification protocol. Extracted DNA was eluted twice for a total volume of 60 μ l, and quantified using a Qubit High Sensitivity fluorometer (Life Technologies) (Supplementary Table S1).

Regional sample set. Samples were prepared and extracted in the paleogenetics clean room at the Institute for Archaeological Sciences, University of Tübingen (INA). The surface of the dedicated sampling hood was cleaned with HPLC water and UV irradiated by an internal light source between uses. Any calculus was removed from the surfaces of the teeth using dental scalers that had been rinsed with bleach and HPLC water and UV irradiated for 10 minutes between uses. Large calculus samples were pulverized with a tube pestle. Teeth were then sectioned horizontally at the cemento-enamel junction and dentin was drilled from the pulp chamber using a dental drill. For calculus samples weighing over 20 mg, half the pulverized material was carried over for extraction. For dentin samples over 70 mg, aliquots of approximately 50 mg were taken for extraction. Dentin and calculus samples were extracted using a previously described silica-based method⁷⁵. In brief, samples were submerged in a digestion buffer with final concentrations of 0.45 M EDTA and 0.25 mg/mL proteinase K (Qiagen, the Netherlands) and rotated at 37 °C until decalcified. After incubation, samples were centrifuged and the supernatant was purified using a 5 M guanidine-hydrochloride binding buffer with High Pure Viral Nucleic Acid Large Volume kits (Roche, Switzerland). The extracts were eluted in 100 μ l of a 10 mM tris-hydrochloride, 1 mM EDTA (pH 8.0), and 0.005% tween-20 buffer (TET). One extraction blank was prepared for every ten samples, and one positive control of cave bear bone powder was processed alongside each extraction batch to ensure efficiency. The extracts were quantified using a Qubit High Sensitivity fluorometer (Life Technologies) (Supplementary Table S1).

Illumina library preparation. *Global sample set.* Approximately 100 ng of DNA was built into each Illumina shotgun library at the LMAMR, Norman, Oklahoma using NEBNext DNA Library Prep Master Set (E6076) and blunt-end modified Illumina adapters. Manufacturer's instructions were followed with the exception of Nebulization. Blunt-end repair was performed using 50 μ l reactions with 30 μ l of DNA extract for each sample which was then incubated for 20 min at 12 °C and 15 min at 37 °C and purified using Qiagen MinElute silica spin columns following the manufacturer's instructions. All samples were eluted in 30 μ l. Following blunt-end repair, Illumina adapters were ligated in 50 μ l reactions. Reactions were incubated for 15 minutes at 20 °C and purified using Qiagen QiaQuick columns before elution in 30 μ l EB. Samples were then incubated for 20 minutes at 37 °C followed by 20 min at 80 °C in a final volume of 50 μ l for adapter fill-in. Libraries were quantified using a real-time quantitative PCR assay (qPCR, Lightcycler 480 Roche) targeting the IS7 and IS8 sequences in the universal Illumina adapters (Supplementary Table S1). Libraries were amplified and dual-indexed in a 50 μ l PCR reaction using 15 μ l template, 25 μ l of a 2 \times KAPA U+ master mix, 5.5 μ l H₂O, 1.5 μ l DMSO, 1 μ l BSA (20 mg/ml), and 1 μ l of each forward and reverse index (10 μ M). Thermocycling conditions were 5 min at 98 °C followed by 10–12 cycles of 15 seconds at 98 °C, 20 seconds at 60 °C, and 20 seconds at 72 °C, followed by a final elongation step for 1 minute at 72 °C. Amplified libraries were then purified using Agencourt AMPure XP beads and eluted in 30 μ l EB. Samples were sent for sequencing on an Illumina HiSeq2500 using a paired-end, 200-cycle, rapid-run chemistry. Samples were sequenced to an average depth of 36 million reads (Supplementary Table S1), and blanks were sequenced to a depth of 0.4–1 million reads (Supplementary Table S10). Deeper sequencing of the blanks was not performed because a high degree of clonality was already achieved at this depth of sequencing.

Regional Sample Set. Double-stranded Illumina libraries were built using 10 μ l of extract for each sample according to an established protocol⁷⁶. Purification of the blunt-end repair and adapter ligation steps was performed using Qiagen MinElute columns. After the adapter fill-in step, the Bst polymerase was deactivated with a 20 minute incubation at 80 °C. A single library blank was used for every ten samples. The libraries were then quantified using real-time qPCR assay (Lightcycler 480 Roche) targeting the IS7 and IS8 sequences of the universal Illumina adapters (Supplementary Table S1). Each library was assigned a unique pair of indices, added to the library over 2–15 indexing reactions per library. Libraries were double-indexed⁷⁷ in 100 μ l reactions using varied amounts of template and H₂O based on library richness, 10 μ l PfuTurbo buffer, 1 μ l PfuTurbo (Agilent Technologies), 1 μ l dNTP mix (25 mM), 1.5 μ l BSA (10 mg/ml), and 2 μ l of each indexing primer (10 μ M). The reactions were purified, pooled, and eluted over MinElute columns in 50 μ l TET. Efficiency of the indexing reactions was evaluated using a qPCR assay. Approximately one-third of each indexed library was amplified using 3–5 μ l of template in 70 μ l reactions with Herculase II Fusion DNA Polymerase (Agilent Technologies). Products for each sample were pooled and quantified using an Agilent Tape Station D1000 Screen Tape kit. Amplified sample and blank libraries were pooled into two 10 nM pools for shotgun sequencing. Samples and blanks were sequenced separately on Illumina NextSeq500 using single-end, 75-cycle, high-output kits. Samples were sequenced to a depth of approximately 5 million reads per library (Supplementary Table S1), and blanks were sequenced to a depth of 157,109 to 841,776 reads (Supplementary Table S10). Additionally, thirteen individuals of the regional sample set were re-sequenced using paired-end 150-cycle chemistry on an Illumina NextSeq500 so they could be included in fragment length analyses.

Computational Methods. Sequencing data from the global and regional sample sets were computationally processed identically using EAGER v1.92⁷⁸. In brief, adapters were removed, paired-end data merged, and reads quality filtered using Clip&Merge⁷⁸ with a minimum base quality of 20 and a minimum fragment length of 30. Processed reads were then taxonomically binned using MALT version 038²⁸ and the full NCBI nucleotide database ('nt', April 2016) with an 85% identity threshold, a minSupportPercent threshold of 0.01, and a topPercent value of 1.0 (Supplementary Tables S2, S3, S14). Metagenomic profiles were analyzed with MEGAN6 CE²⁹ and screened for specific taxonomic levels for fragment length and damage pattern profiles using a RMA file format

parsing script⁷⁹. Length and damage patterns were visualized using the ggplot2⁸⁰ and lattice⁸¹ libraries in R. For all fragment length and damage rate analyses only those samples that were paired end sequenced were used. Additionally, only merged reads were considered to prevent artifacts brought on by unmerged pairs, the maximum length of which is fixed. Mapping to the human genome (hg19) was performed using BWA ($-n\ 0.01, -l\ 1000, -q\ 30$)⁸² as implemented in EAGER v1.92⁷⁸. A normalized taxon table with species-level assigned reads (Supplementary Table S2) was exported from the bacterial and archaeal sub-trees in MEGAN and used to generate a Bray-Curtis taxonomic distance matrix in R using the vegan library (version 2.4-1)⁸³. Principal coordinates were generated using the R ape library⁸⁴ and visualized as a PCoA plot using ggplot2⁸⁰. The same taxon table was used to produce microbial profiles of each sample according to isolation source within a nested classification scheme⁸⁵. This scheme illustrates the relative environmental and human microbial contributions to each sample based on a species-by-species literature survey based on species name searches in PubMed. Pathobionts and opportunistic pathogens are designated as such when literature on the organism consistently presented it as a health threat, though it also may be a natural inhabitant of the human microbiome or soil. Assigned read counts for the classified species were tabulated and visualized using the Krona Excel Template⁸¹. Potential source contribution to samples was also calculated from summarized read quantities of the bacterial and archaeal sub-trees (Supplementary Table S3) using SourceTracker version 0.9.8³⁴. Source accession numbers: ERR1017187, ERR1019366, ERR1022687, ERR1022692, ERR1034454, ERR1035437, ERR1035438, ERR1035441, ERR1039457, ERR1039458, ERR1043165, ERR1044071, ERR1044072, ERR1051325, SRR1631060, SRR1631061, SRR1631063, SRR1631064, SRR1633008, SRR3184100, SRR3184876, SRR3189411, SRR3189416, SRR3189418, SRS014107, SRS015650, SRS018665, SRS018975, SRS019029, SRS019129, SRS019387, SRS023538, SRS063215, SRS077312. The microbial classification scheme with Krona plots and SourceTracker were used in conjunction to control for potential biases related to the modern samples used in the latter and together present a layered representation of the types of microorganisms typically found in ancient dental calculus and dentin. Finally, GC content versus length profiles were generated using a custom python script (<https://github.com/aemann01/gcLenCorPlots>). When possible, analyses were run in parallel using GNU parallel⁸⁶.

Accession numbers. Genetic data have been deposited in the NCBI Short Read Archive (SRA) under the Bioproject accession PRJNA445215.

References

- White, D. J. Processes contributing to the formation of dental calculus. *Biofouling* **4**, 209–218 (1991).
- Marsh, P. D. Dental plaque as a biofilm and microbial community - implications for health and disease. *BMC Oral Health* **6**, <https://doi.org/10.1186/1472-6831-6-S1-S14> (2006).
- Warinner, C., Speller, C. & Collins, M. J. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Phil. Trans. R. Soc. B* **370**, 20130376 (2015).
- Marsh, P. D. & Bradshaw, D. J. Dental plaque as a biofilm. *Journal of Industrial Microbiology* **15**, 169–175, <https://doi.org/10.1007/bf01569822> (1995).
- Power, R. C., Salazar-García, D. C., Wittig, R. M., Freiburger, M. & Henry, A. G. Dental calculus evidence of plant diet and life history transitions in the chimpanzees of the Tai Forest. *Scientific Reports* **5**, 15161 (2015).
- Power, R. C., Salazar-García, D. C., Wittig, R. M. & Henry, A. G. Assessing use and suitability of scanning electron microscopy in the analysis of micro remains in dental calculus. *Journal of Archaeological Science* **49**, 160–169 (2014).
- Armitage, P. L. The extraction and identification of opal phytoliths from the teeth of ungulates. *Journal of Archaeological Science* **2**, 187–197, [https://doi.org/10.1016/0305-4403\(75\)90056-4](https://doi.org/10.1016/0305-4403(75)90056-4) (1975).
- Fox, C. L., Pérez-Pérez, A. & Juan, J. Dietary Information through the Examination of Plant Phytoliths on the Enamel Surface of Human Dentition. *Journal of Archaeological Science* **21**, 29–34, <https://doi.org/10.1006/jasc.1994.1005> (1994).
- Henry, A. G., Brooks, A. S. & Piperno, D. R. Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium). *Proceedings of the National Academy of Sciences* **108**, 486 (2011).
- Dudgeon John, V. & Tromp, M. Diet, Geography and Drinking Water in Polynesia: Microfossil Research from Archaeological Human Dental Calculus, Rapa Nui (Easter Island). *International Journal of Osteoarchaeology* **24**, 634–648, <https://doi.org/10.1002/oa.2249> (2012).
- King, D. J., Searcy, M. T., Yost, C. L. & Waller, K. Corn, Beer, and Marine Resources at Casas Grandes, Mexico: An Analysis of Prehistoric Diets Using Microfossils Recovered from Dental Calculus. *Journal of Archaeological Science: Reports* **16**, 365–379, <https://doi.org/10.1016/j.jasrep.2017.10.013> (2017).
- de la Fuente, C., Flores, S. & Moraga, M. DNA from human ancient bacteria: A novel source of genetic evidence from archaeological dental calculus. *Archaeometry* **55**, 767–778 (2013).
- Adler, C. J. *et al.* Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat Genet* **45**(450–455), 455e451, <https://doi.org/10.1038/ng.2536> (2013).
- Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344 (2014).
- Weyrich, L. S. *et al.* Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**, 357 (2017).
- Price, S. D. R., Keenleyside, A. & Schwarcz, H. P. Testing the validity of stable isotope analyses of dental calculus as a proxy in paleodietary studies. *Journal of Archaeological Science* **91**, 92–103 (2018).
- Ozga, A. T. *et al.* Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *American journal of physical anthropology* **160**, 220–228 (2016).
- Paabo, S. *et al.* Genetic analyses from ancient DNA. *Annual review of genetics* **38**, 645–679, <https://doi.org/10.1146/annurev.genet.37.110801.143214> (2004).
- Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA damage. *Cold Spring Harbor perspectives in biology* **5**, a012567 (2013).
- Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic acids research* **29**, 4793–4799 (2001).
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).
- Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports* **5**, 11184 (2015).
- Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* **5**, 5257, <https://www.nature.com/articles/ncomms6257-supplementary-information>, <https://doi.org/10.1038/ncomms6257> (2014).
- Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS one* **10**, e0129102 (2015).
- Hansen, H. B. *et al.* Comparing ancient DNA preservation in petrous bone and tooth cementum. *PLoS One* **12**, e0170940 (2017).

26. Ziesemer, K. A. *et al.* Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Sci Rep* **5**, 16498, <https://doi.org/10.1038/srep16498> (2015).
27. Kemp, B. M., Winters, M., Monroe, C. & Barta, J. L. How much DNA is lost? Measuring DNA loss of short-tandem-repeat length fragments targeted by the PowerPlex 16 System using the Qiagen MinElute Purification Kit. *Human Biology* **86**, 1–18 (2014).
28. Vagene, A. J. *et al.* Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, <https://doi.org/10.1038/s41559-017-0446-6> (2018).
29. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology* **12**, e1004957, <https://doi.org/10.1371/journal.pcbi.1004957> (2016).
30. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biology* **12**, R60 (2011).
31. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**, 385 (2011).
32. Chen, T. *et al.* The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* **2010** (2010).
33. Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46**, D8–D13, <https://doi.org/10.1093/nar/gkx1095> (2018).
34. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat Meth* **8**, 761–763, <http://www.nature.com/nmeth/journal/v8/n9/abs/nmeth.1650.html> - supplementary-information (2011).
35. Méthé, B. A. *et al.* A framework for human microbiome research. *Nature* **486**, 215 (2012).
36. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207 (2012).
37. Oh, J. *et al.* Temporal stability of the human skin microbiome. *Cell* **165**, 854–866 (2016).
38. Johnston, E. R. *et al.* Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. *Frontiers in microbiology* **7**, 579 (2016).
39. Belda-Ferre, P. *et al.* The oral metagenome in health and disease. *The ISME journal* **6**, 46 (2012).
40. Jiang, W. *et al.* Pyrosequencing analysis of oral microbiota shifting in various caries states in childhood. *Microbial ecology* **67**, 962–969 (2014).
41. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684, <https://doi.org/10.1093/bioinformatics/btt193> (2013).
42. Schuenemann, V. J. *et al.* Genome-Wide Comparison of Medieval and Modern Mycobacterium leprae. *Science* **341**, 179 (2013).
43. Socransky, S. S. & Haffajee, A. D. Dental biofilms: difficult therapeutic targets. *Periodontology 2000* **28**, 2–55 (2002).
44. Marshall, G. W., Marshall, S. J., Kinney, J. H. & Balooch, M. The dentin substrate: structure and properties related to bonding. *Journal of Dentistry* **25**, 441–458, [https://doi.org/10.1016/S0300-5712\(96\)00065-6](https://doi.org/10.1016/S0300-5712(96)00065-6) (1997).
45. Palmer, L., Chapple, I., Wright, H., Roberts, A. & Cooper, P. Extracellular deoxyribonuclease production by periodontal bacteria. *Journal of periodontal research* **47**, 439–445 (2012).
46. Warinner, C., Speller, C., Collins, M. J. & C. M., L. Jr. Ancient human microbiomes. *Journal of Human Evolution* **79**, 125–136 (2015).
47. SHIRATO, M. *et al.* Observations of the surface of dental calculus using scanning electron microscopy. *The Journal of Nihon University School of Dentistry* **23**, 179–187 (1981).
48. Watson, P. *et al.* Penetration of fluoride into natural plaque biofilms. *Journal of dental research* **84**, 451–455 (2005).
49. Velsko, I. M. *et al.* The dental calculus metabolome in modern and historic samples. *Metabolomics* **13** (2017).
50. Linde, A. & Goldberg, M. Dentinogenesis. *Critical Reviews in Oral Biology and Medicine* **4**, 679–728 (1993).
51. Lekic, P., Rojas, J., Birek, C., Tenenbaum, H. & McCulloch, C. A. G. Phenotypic comparison of periodontal ligament cells *in vivo* and *in vitro*. *Journal of Periodontal Research* **36**, 71–79 (2001).
52. Provenza, D. V. The Blood Vascular Supply of the Dental Pulp with Emphasis on Capillary Circulation. *Circulation Research* **6**, 213 (1958).
53. Phillips, A. *et al.* Comprehensive analysis of microorganisms accompanying human archaeological remains. *GigaScience* (2017).
54. Fuchs, T. A. *et al.* Novel cell death program leads to neutrophil extracellular traps. *Journal of Cell Biology* **176**, 231–241 (2007).
55. Seper, A. *et al.* Vibrio cholerae evades neutrophil extracellular traps by the activity of two extracellular nucleases. *PLoS pathogens* **9**, e1003614 (2013).
56. Ryder, M. I. Comparison of neutrophil functions in aggressive and chronic periodontitis. *Periodontology 2000* **53**, 124–137 (2010).
57. Branzk, N. *et al.* Neutrophils sense microbe size and selectively release neutrophil extracellular traps in response to large pathogens. *Nature immunology* **15**, 1017 (2014).
58. von Köckritz-Blickwede, M., Blodkamp, S. & Nizet, V. Interaction of bacterial exotoxins with neutrophil extracellular traps: impact for the infected host. *Frontiers in microbiology* **7**, 402 (2016).
59. Bhongir, R. K. *et al.* DNA-fragmentation is a source of bactericidal activity against Pseudomonas aeruginosa. *Biochemical Journal* **474**, 411–425 (2017).
60. Majewski, P. *et al.* Inhibitors of serine proteases in regulating the production and function of neutrophil extracellular traps. *Frontiers in immunology* **7**, 261 (2016).
61. Brinkmann, V. & Zychlinsky, A. Neutrophil extracellular traps: is immunity the second function of chromatin? *J Cell Biol* **198**, 773–783 (2012).
62. Palmer, L. J., Damgaard, C., Holmstrup, P. & Nielsen, C. H. Influence of complement on neutrophil extracellular trap release induced by bacteria. *Journal of periodontal research* **51**, 70–76 (2016).
63. Dang, G. *et al.* Characterization of Rv0888, a novel extracellular nuclease from Mycobacterium tuberculosis. *Scientific reports* **6**, 19033 (2016).
64. Warinner, C. *et al.* A Robust Framework for Microbial Archaeology. *Annual Review of Genomics and Human Genetics* **18**, 321–356, <https://doi.org/10.1146/annurev-genom-091416-035526> (2017).
65. Panjkovich, A., Norambuena, T. & Melo, F. dnaMATE: a consensus melting temperature prediction server for short DNA sequences. *Nucleic Acids Research* **33**, W570–W572, <https://doi.org/10.1093/nar/gki379> (2005).
66. Foerster, K. U., Von Mering, C., Hooper, S. D. & Bork, P. Environments shape the nucleotide composition of genomes. *EMBO reports* **6**, 1208–1213 (2005).
67. Gansauge, M.-T. *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research* **45**, e79–e79, <https://doi.org/10.1093/nar/gkx033> (2017).
68. Waters-Rist, A. L. & Hoogland, M. L. P. Osteological evidence of short-limbed dwarfism in a nineteenth century Dutch family: Achondroplasia or hypochondroplasia. *Int J Pathology* **3**, 243–256 (2013).
69. Maurandi, J. L., Martinex, M. L., Martinez, F. R. & Fernandez, A. A. The collective Chalcolithic burial of Camino del Molino (Caravaca de la Cruz, Murcia, Spain). Methodology and the first results of an exceptional archaeological site. *Trabajos Prehist* **66**, 143–159 (2009).
70. Aldenderfer, M. Variation in mortuary practice on the early Tibetan plateau and the high Himalayas. *Journal of the International Association for Bone Research* **1**, 293–318 (2013).
71. Jeong, C. *et al.* Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci USA* **113**, 7485–7490, <https://doi.org/10.1073/pnas.1520844113> (2016).
72. Littleton, J. *et al.* Taphonomic analysis of Bronze Age burials in Mongolian khirigsuurs. *Journal of Archaeological Science* **39**, 3361–3370 (2012).

73. Stone, A. C. & Stoneking, M. Ancient DNA from a pre-columbian Amerindian population. *American Journal of Physical Anthropology* **92**, 463–471 (1993).
74. Read, C. In *Medieval Lough Cé: history, archaeology, and landscape* (ed. Thomas Finan) (Four Courts Press, 2010).
75. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* **110**, 15758–15763 (2013).
76. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb. prot5448 (2010).
77. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* **40**, e3 (2012).
78. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome biology* **17**, 60 (2016).
79. Hübner, R. *et al.* In *ISMB/ECCB* (Prague, Czech Republic, 2017).
80. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2009).
81. Sarkar, D. *Lattice: multivariate data visualization with R*. (Springer Science & Business Media, 2008).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. Oksanen, J. *et al.* *vegan: Community Ecology Package*. <https://CRAN.R-project.org/package=vegan> (2016).
84. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
85. Sabin, S. *Ancient DNA analysis of dental remains from Kiltasheen: a case study in metagenomics and an exploration of dental calculus*. M.Sc. thesis, Eberhard-Karls-Universität Tübingen, (2016).
86. Tange, O. In *The USENIX Magazine* 42-47 (2011).

Acknowledgements

This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) ERC grant agreement (319209 'Nexus 1492' to C.H.) and the US National Science Foundation (BCS-1516633 and BCS-1528698 to C.W., BCS-1515163 to A.C.S., and BCS-1643318 to C.W. and C.A.H.). Funding for open access charge: Max Planck Society. The authors also thank Ron Hübner and Felix Key for access to their RMA parsing script. Thanks to Joaquín Lomba, María Haber and Azucena Avilés for allowing sampling of the Camino del Molino samples.

Author Contributions

C.W. and K.I.B. designed the study. A.M., S.S., K.Z., Á.J.V., H.S., A.T.O., C.A.H. and D.S.G. performed the experiments. A.M., S.S., K.S., and J.F.Y. analyzed the data. D.S.G., B.F., M.A., M.H., C.R., G.M., A.C.S., C.M.L., J.K. and C.H. provided materials and resources. A.M., S.S., and C.W. wrote the manuscript with input from all other authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-28091-9>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

SUPPLEMENTARY INFORMATION

Differential preservation of endogenous human and microbial DNA in dental calculus and dentin

Allison E. Mann, Susanna Sabin, Kirsten Ziesemer, Åshild J. Vågane, Hannes Schroeder, Andrew T. Ozga, Krithivasan Sankaranarayanan, Courtney A. Hofman, James A. Fellows Yates, Domingo C. Salazar-Garcia, Bruno Frohlich, Mark Aldenderfer, Menno Hoogland, Christopher Read, George R. Milner, Anne C. Stone, Cecil M. Lewis, Jr., Johannes Krause, Corinne Hofman, Kirsten Bos, Christina Warinner

Correspondence to: warinner@shh.mpg.de

Supplementary Methods

Linear Discriminant Analysis (LDA) of Microbial Communities

MALT results for dentin, calculus, and blank samples (excluding NF47 and NF217 samples) were extracted as a taxon table representing the whole bacterial and archaeal sub-trees, with taxon abundances expressed as normalized, summarized read counts. Samples were classified by material type (dentin, n=45; calculus, n=47; blank, n=33). Any taxa that were not present in at least 30% of the samples in at least one class were excluded from the analysis. We conducted an LDA on the resulting taxon table with the LDA effect size (LEfSe) method, available on the Galaxy server (Segata et al. 2011). The alpha value for the Kruskal-Wallis test among classes was set to 0.05, and the minimum logarithmic LDA score threshold was set to 2.0.

Human Read Validation:

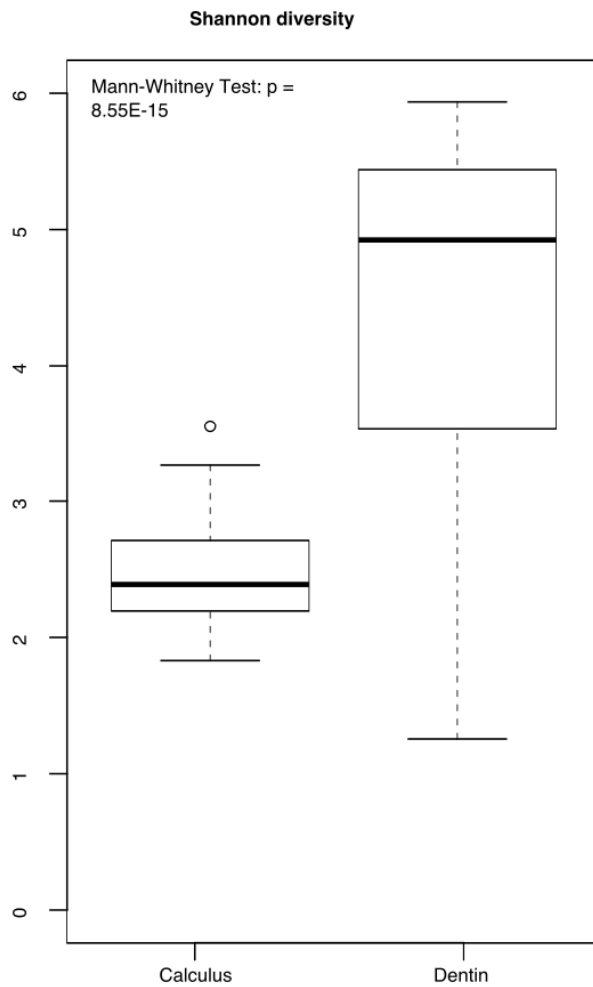
Reads mapped to the hg19 human reference genome using sensitive BWA (Li and Durbin 2009) mapping parameters (-n 0.01, -l 1000, -q 30) were de-replicated using DeDup as implemented in EAGER version 1.92 (Peltzer, Jäger et al. 2016). Next, de-replicated bam files were converted to fastq format using bedtools version 2.25.0-1 (bedtools bamtofastq) (Quinlan 2010) and where necessary split into forward, reverse, and merged reads. Three base pairs were trimmed from either side of all merged reads using fastx_trimmer from the FASTX-Toolkit version 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Forward reads were trimmed of three base pairs exclusively on the 5' end of the read while reverse pair reads were trimmed for three base pairs on the 3' end. Trimmed reads were then remapped to the hg19 human reference genome using BWA with a higher mismatch penalty and quality mapping threshold so that approximately one mismatch would be allowed per 50 bases (-n 0.2, -l 1000, -q 37). The reads passing this quality threshold filter have a higher level of confidence of their proper assignment to the human genome. To further test the validity of these reads, they were then run through a lowest-common-ancestor algorithm via MALT version 038 using the full NCBI NT database with a percent identity threshold of 90%. Those reads that were assigned to the *Homo sapiens* node are recorded in Supplementary Table S9. Finally, to test whether these high confidence human reads are in fact ancient and not the result of background contamination, the original reads pre-damage trimming were pulled from the original fastq files, mapped to the hg19 human reference genome using BWA (-n 0.01, -l 1000, -q 30) in the EAGER v1.92 pipeline (Peltzer, Jäger et al. 2016) and assessed for terminal cytosine deamination patterns using mapDamage version 2.0 (Jonsson, Ginolhac et al. 2013). For all calculus samples and most dentin samples, the percent endogenous and damage patterns both pre- and post-strict map filtering are comparable, confirming the observed pattern of a low but consistent human aDNA retrieval from calculus and a variable yield from dentin (Supplementary Figure S5).

As an additional assessment of exogenous contamination, mitochondrial contamination estimates were generated using Schmutzi for all samples demonstrating sufficient (i.e. at least five-fold) coverage (Renaud et al. 2015) of the human mitochondrial reference genome (Supplementary Table S10).

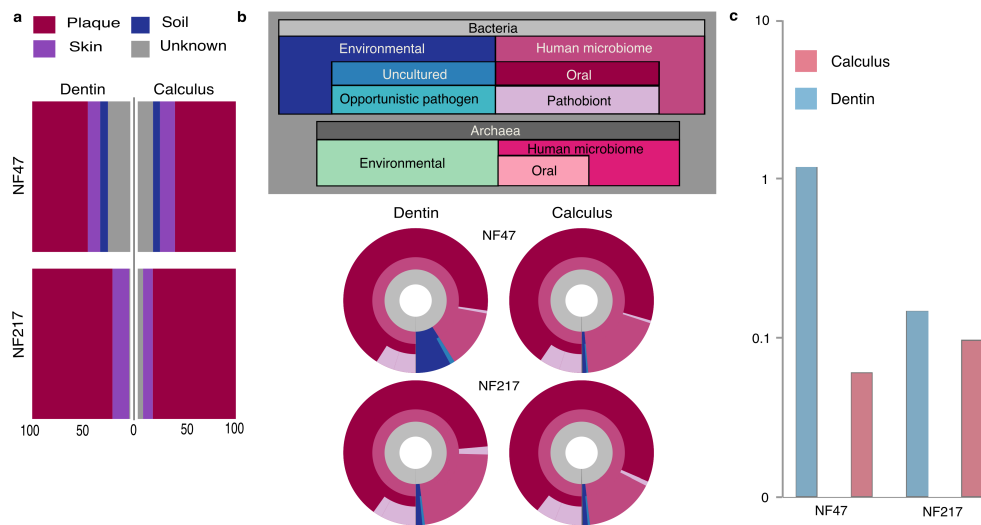
Bacterial fragment analyses:

Reads that mapped to the species node and all higher resolution taxonomic nodes were extracted from the MALT results for all bacteria of interest. To limit the impact of erroneous mapping only those reads with damage at the terminus of the read were considered. For all fragment length analyses, only merged reads were analyzed.

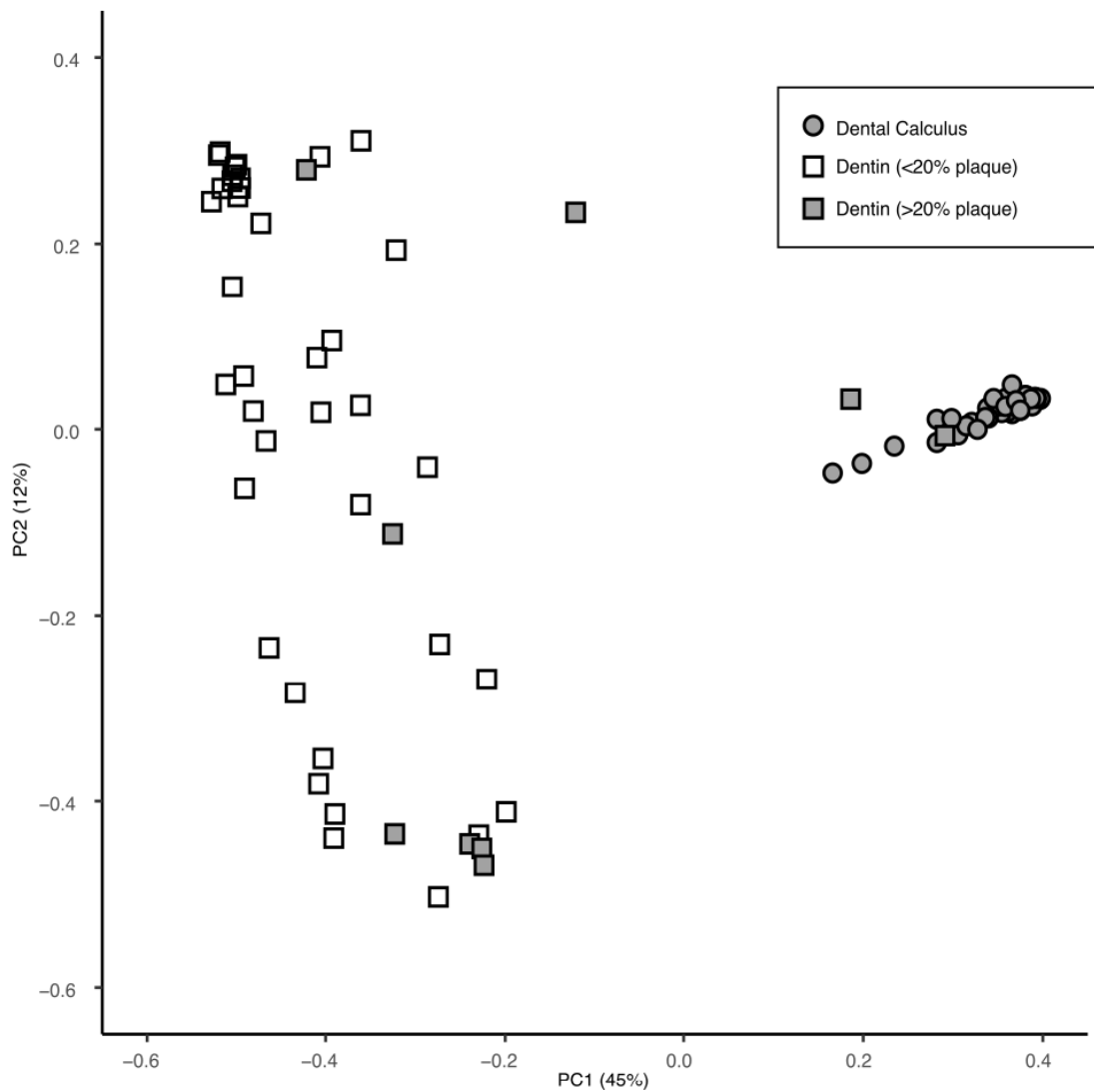
Supplementary Figures



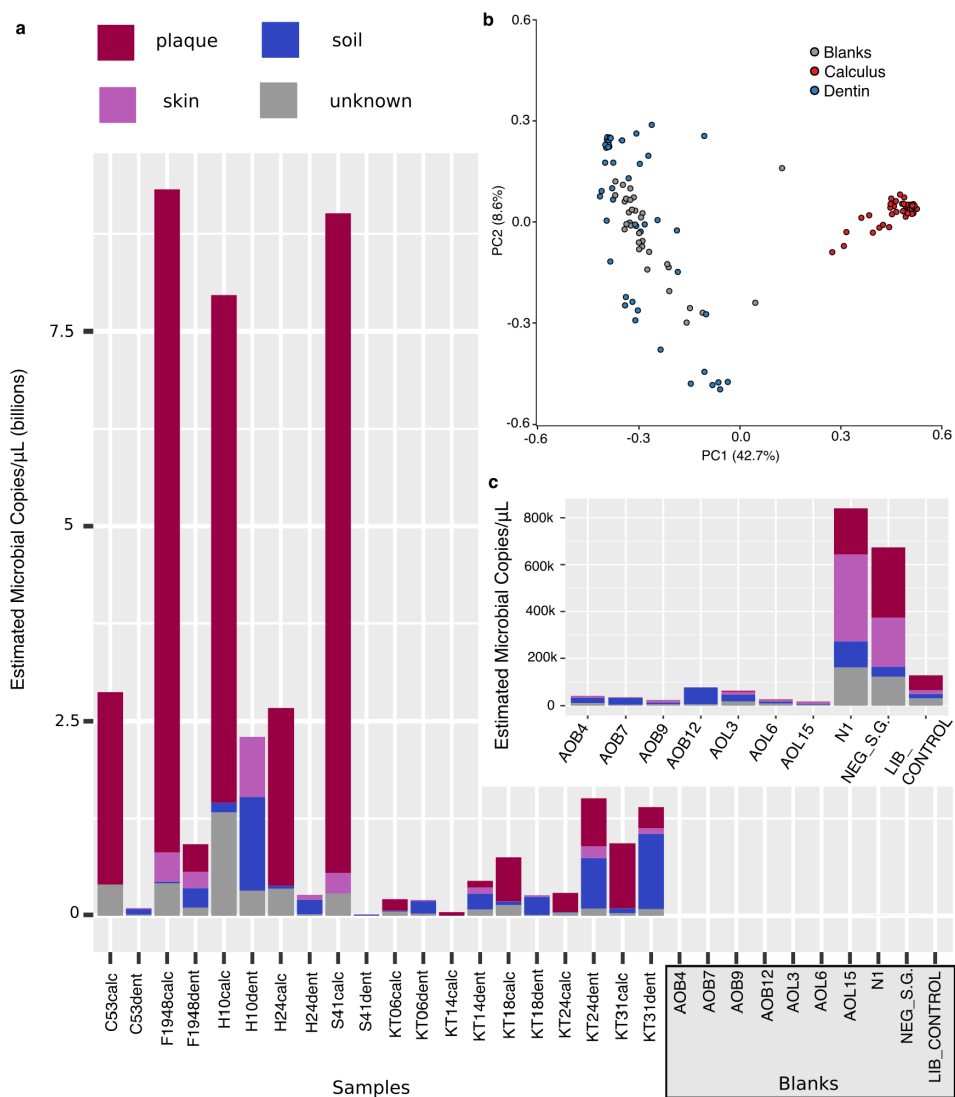
Supplementary Figure S1: Boxplots of alpha diversity (Shannon Weaver Index) for all dentin and dental calculus samples. Diversity among dental calculus is significantly lower (Mann-Whitney-U $p = 8.55E-15$) as compared to that found in the dentin samples.



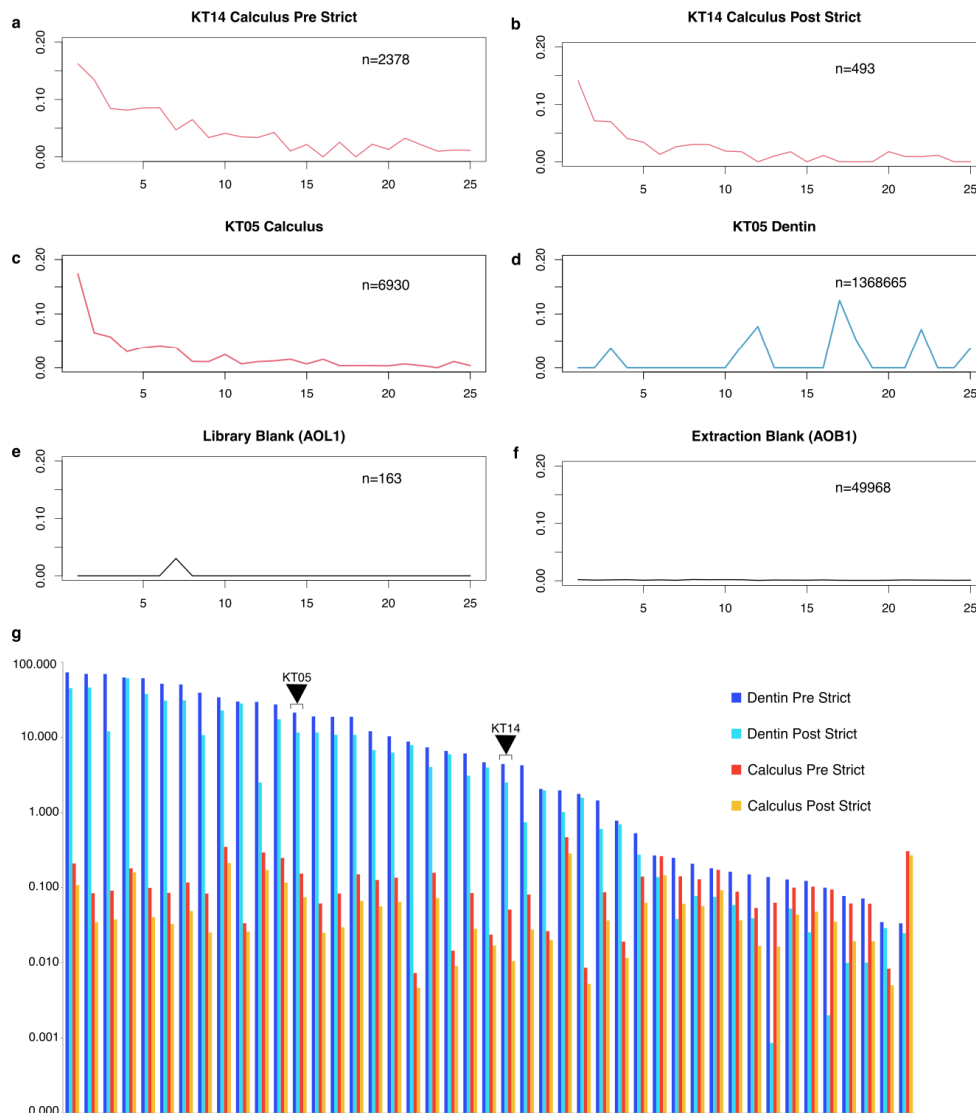
Supplementary Figure S2: Likely signal of carious lesions on two dentin samples. (a) Bar chart displaying Bayesian SourceTracker results of the estimated genetic contributions from human oral and environmental microbial sources. (b) The donut plots are constructed from nested classification of species-level MALT results. The microbial communities in the Norris Farms dentin and calculus samples have a high oral contribution, as demonstrated by both the SourceTracker bar plots and the MALT classification donut plots. (c) Log transformed (\log_{10}) proportion human endogenous DNA in NF47 and NF217 dentin and calculus samples reflecting a relatively high proportion of microbial DNA in the Norris Farms dentin samples.



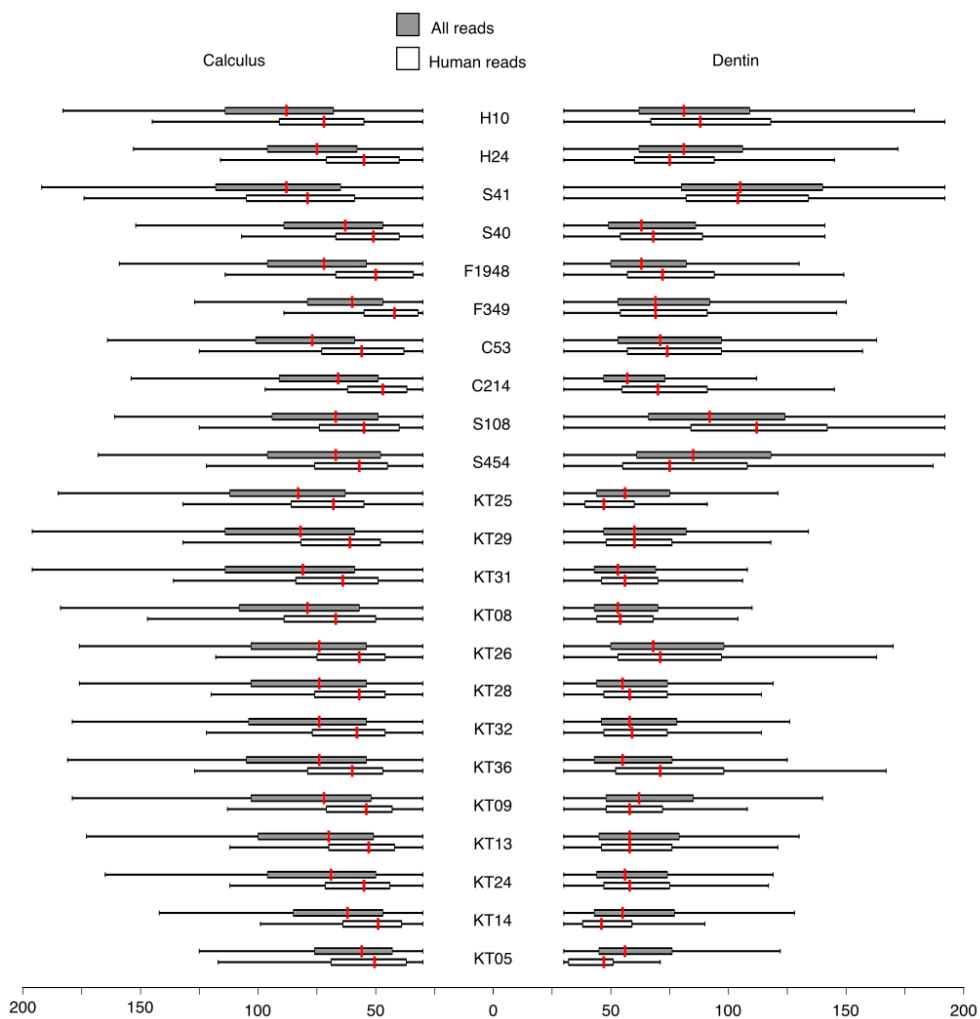
Supplementary Figure S3. Principal Coordinates Analysis (PCoA) of Bray-Curtis distances of all bacterial and archaeal species-level assignments from dental calculus and dentin. Dental calculus is represented as circles and dentin is represented as squares. Color indicates estimated contribution of oral taxa. Black symbols are those samples that have a predicted proportion of oral contribution of 20% or more, illustrating that some dentin samples have an unexpectedly high oral signature, though most do not cluster with the dental calculus samples, indicative of a non-biological community.



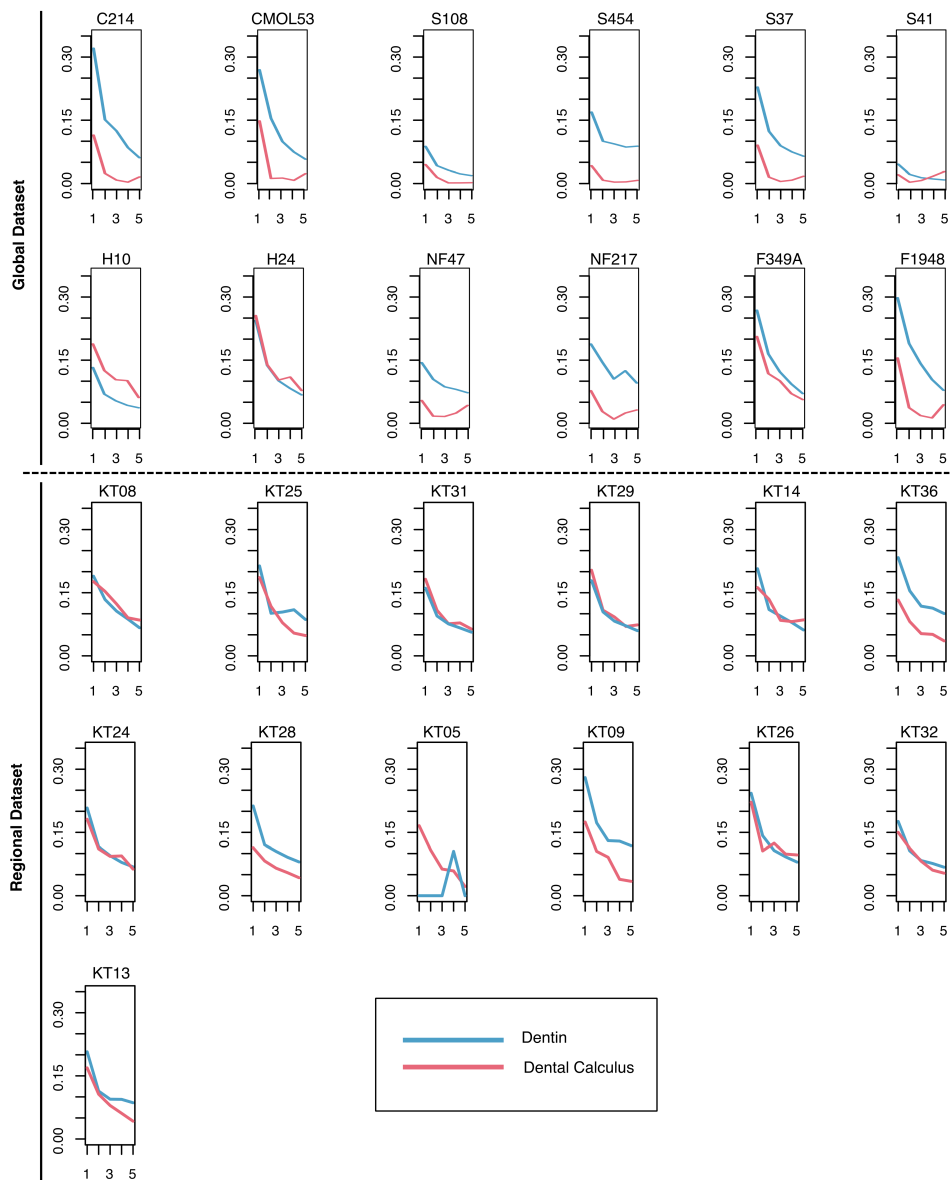
Supplementary Figure S4. Comparison of microbial DNA quantity and composition between samples and blanks. (a) Estimated number of copies of microbial DNA per microliter of pre-indexed library, with proportion of predicted sources indicated. Predicted sources of microbial DNA were calculated for each sample using SourceTracker. Total microbial copies/ μ L was calculated for each sample by subtracting the proportion of endogenous human DNA (Supplementary Tables S9 and S10) from the total copies/ μ L (Supplementary Table S1). **(b)** PCoA of Bray-Curtis dissimilarity generated from a normalized taxon table of bacterial and archaeal species-level abundances for all samples (excluding NF47 and NF217) and blanks. **(c)** Enhanced view of the estimated source proportions and number of copies of microbial DNA per microliter of pre-indexed blank libraries; enhanced view is shown because the quantity of DNA measured in the blanks was too low to be visible in panel **(a)**.



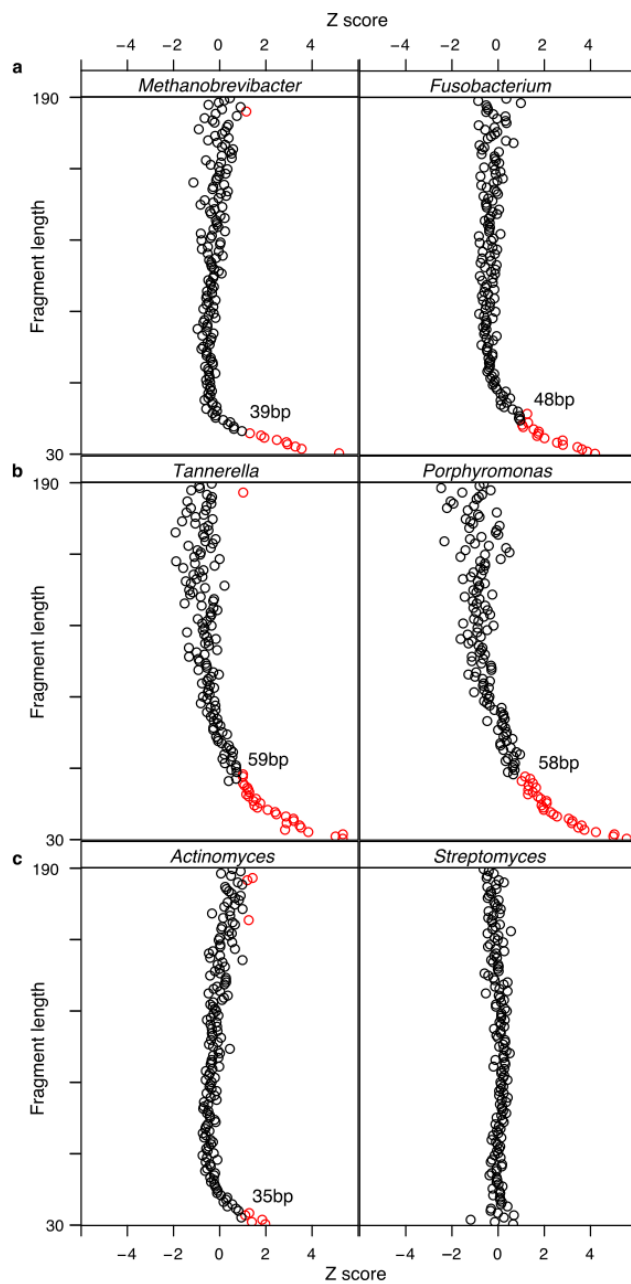
Supplementary Figure S5. Validation of ancient human DNA authenticity. (a) Terminal cytosine deamination rates of reads mapped to the human genome from a single calculus sample (KT14) before extra human validation steps. (b) Damage rate of human reads from calculus sample KT14 after human validation. While the damage rate of human reads post-validation drops, a clear damage signal, consistent with authentic ancient DNA is still observed. (c-d) Terminal cytosine deamination rates of reads mapped to the human genome from a calculus and dentin pair where the calculus exhibits a clear damage signal consistent with authentic ancient DNA while the dentin sample is consistent with modern contamination. (e-f) Terminal cytosine deamination of a library blank and extraction blank. Neither blanks have an observed damage signal, consistent with modern DNA. (g) Proportion human endogenous content for all paired samples both before and after strict mapping (see Supplementary Methods). While in all cases the proportion of human endogenous content drops after strict mapping, the effect is minor for most samples. Major drops are detected in certain dentin samples.



Supplementary Figure S6: Boxplots comparing fragment size distributions of human and non-human reads in each sample. For calculus, human reads are consistently shorter than those that map to microbial sources. By contrast, no consistent pattern is observed for dentin samples.



Supplementary Figure S7. Differences in microbial damage patterns among paired dentin and dental calculus is sample-specific. Cytosine damage patterns for a subset of paired dentin (blue) and dental calculus (red) samples. While most dental calculus samples have a lower initial deamination rate than their dentin pair in the global dataset, this pattern is not consistently observed in the regional dataset, possibly the result of differences in laboratory preparation.



Supplementary Figure S8. Fragment length deviation from expected mean GC content for selected bacterial genera. Bacterial genera are organized by expected genomic GC content wherein (a) are low genomic GC taxa, (b) are moderate genomic GC taxa and, (c) are high genomic GC taxa. Each point represents a single length bin's mean deviation from the overall mean of all reads mapped to the genus. Red points are those length bins that are one or more Z scores deviated from the mean GC content. For each genus the read length bin at which a major deviation can be seen (± 1 z score) is noted on the graph. For low or medium genomic GC content genera, this length threshold occurs at a higher fragment length than those with high genomic GC content.

Supplementary References

- Jonsson, H., A. Ginolhac, M. Schubert, P. L. Johnson and L. Orlando (2013). "mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters." *Bioinformatics* **29**(13): 1682-1684.
- Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler Transform." *Bioinformatics* **25**: 1754-1760.
- Peltzer, A., G. Jäger, A. Herbig, A. Seitz, C. Kniep, J. Krause and K. Nieselt (2016). "EAGER: efficient ancient genome reconstruction." *Genome Biology* **17**(1): 60.
- Quinlan, A. (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* **26**(6): 841-842.
- Renaud, Gabriel et al. (2015). "Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA." *Genome Biology* **16**: 224
- Segata, N., et al. (2011). "Metagenomic biomarker discovery and explanation." *Genome Biology* **12**(6): R60

6. Manuscript C

Microbial preservation and population health inferred from metagenomic analyses of two medieval latrines

Susanna Sabin, Marissa Ledger, Hui-Yuan Yeh, Piers Mitchell, Kirsten I. Bos

ABSTRACT

Ancient latrine sediments, containing the concentrated collective solid waste of whole human communities, are theoretically ideal proxies for human health on the group level. They have been heavily utilized in the past to detect the presence of gut-associated eukaryotic parasites through microscopy, immunoassays, and genetics. However, there have been no exploratory studies to date investigating the whole genetic content of ancient latrine sediments including not only the presence of gastrointestinal parasites, but also the core community gut microbiome of the group that used the latrine. Here, we present results from the metagenomic analysis of bulk sediment from medieval latrines in Riga, Latvia and Jerusalem. Our findings supplement and strengthen prior microscopic and chemical analyses of these sediments, particularly regarding the presence of *Ascaris lumbricoides* at both sites. Our findings also highlight the ways investigations of eukaryotic parasites in past human populations will improve with the expansion of publicly available genomic references for many species.

INTRODUCTION

Studies of ancient parasites and ancient human gut microbiota in paleofeces and latrine sediments have provided a vivid window into past population health, dietary practices, and movement. In the past, eukaryotic parasites have been identified through microscopy (Fugassa et al., 2010; Han et al., 2003; Mitchell et al., 2011; Reinhard et al., 1987; Yeh et al., 2015, 2014), immunoassays (Gonçalves et al., 2002; Le Bailly et al., 2008; Mitchell et al., 2008; Yeh et al., 2015, 2014), and genetics (Côté et al., 2016; Guhl et al., 1999; Iñiguez et al., 2006, 2003; Loreille et al., 2001; Maicher et al., 2017; Oh et al., 2010b, 2010b; Søe et al., 2018). These studies have demonstrated the antiquity of many eukaryotic parasites in human populations, and suggest co-evolutionary relationships in many cases (Araújo et al., 2013, 2008; Araújo and Ferreira, 2000; Mitchell, 2013; Perry, 2014). Meanwhile, investigations of the gut microbiome in paleofeces are limited to genetics, and comparatively few papers have been released on this subject (Cano et al., 2014; Santiago-Rodriguez et al., 2015). Genetic investigation of latrine sediments offers the best avenue to analyze these human-associated microbes – commensal gut microbes and pathogenic eukaryotic parasites – in tandem.

Here, we join ancient DNA with previously published paleoparasitology investigations to elucidate human health at two archaeological sites in Jerusalem and Riga, Latvia.

Sediments from each site were microscopically screened in an earlier work for parasite eggs and tested by Enzyme Linked Immunosorbent Assay (ELISA) for specific species of protozoan eukaryotic parasite (Yeh et al., 2015, 2014), and the findings supported the sites' reputations as hubs for travel and trade. Samples from these sites were provided for metagenomic screening to generally explore the genetic composition of the samples and more specifically seek to support and supplement the results of previous microscopy and ELISA analyses. The Jerusalem latrine (MAM) was located in the Christian Quarter, and archaeological and radiocarbon dating suggest a period of use during the fifteenth century. Microscopic analyses of disaggregated sediment and twelve individual coprolites revealed eggs from *Trichuris trichiura* (whipworm), *Ascaris lumbricoides* (roundworm), *Taenia* sp. (tapeworm), and *Diphyllobothrium* sp. (fish tapeworm). The protozoans *Entamoeba histolytica* and *Giardia duodenalis* were identified by ELISA. The latrine site in Riga was discovered in the medieval "Liv quarter," and dated to 1356 CE by dendrochronology. Eggs belonging to *Trichuris* sp., *Diphyllobothrium* sp., *Ascaris* sp., and *Oxyuris equi* (equid pinworm) were identified microscopically in the Riga sediments, and *Entamoeba histolytica* was identified by ELISA analysis.

Here, we present metagenomic analyses of bulk latrine sediments from both sites. We report library inhibition and the results of subsequent template reduction, alpha diversity of the total metagenome from each site, estimated source contributions, and the preservation of eukaryotic parasite DNA. We discuss similarities and differences between the findings from the original paleoparasitological analyses and those from the genetic analysis, including the benefits and limitations of our methodology.

RESULTS

Library inhibition

Two sediment samples from the Jerusalem (MAM) and Riga (RGA) latrine sites were extracted. The eluate of the MAM extracts was observed to have a brown tinge, and subsequently we decided to build one set of libraries for MAM using 10 μ l of DNA extract (as described in our standard library preparation protocol) and one set using 2 μ l of DNA extract. For extracts without inhibitory chemicals disabling the action of the polymerase in the quantitative polymerase chain reaction (qPCR) assay (Wilson, 1997), we would expect the libraries built with 10 μ l of extract to contain approximately five times as many DNA fragments as the libraries built with 2 μ l of extract, as measured by qPCR after ligation and fill-in of Illumina adapters, but prior to indexing and amplification. However, we found that all MAM libraries, regardless of DNA input volume, had approximately the same number of DNA fragments (Table 1). This suggested the libraries were inhibited, and we only took the low template volume MAM libraries forward for further processing and analysis. We later produced low template libraries for RGA as well, though the discoloration of those extracts

was not as extreme. In the case of RGA, the low template libraries were both richer in DNA fragments than the standard template libraries (Table 1). To replicate the complexity of a standard 10 μ L template library, we generated five libraries with 2 μ L of template from each MAM extract, and combined them after indexing. Their pre-indexing quantitation yielded expected values.

Library ID	Library Type	Quant1	Quant2	Mean Copies/ μ L
MAM001.A1	10 μ L template (standard)	1.22E+07	1.25E+07	1.24E+07
MAM001.A2	2 μ L template	1.89E+07	1.78E+07	1.84E+07
MAM001.A3-1	2 μ L template*	2.02E+07	1.95E+07	1.99E+07
MAM001.A3-2	2 μ L template*	1.26E+07	1.48E+07	1.37E+07
MAM001.A3-3	2 μ L template*	1.95E+07	2.10E+07	2.03E+07
MAM001.A3-4	2 μ L template*	1.77E+07	1.95E+07	1.86E+07
MAM001.A3-5	2 μ L template*	2.05E+07	2.32E+07	2.19E+07
MAM001.B1	10 μ L template (standard)	3.26E+07	3.37E+07	3.32E+07
MAM001.B2	2 μ L template	4.13E+07	5.10E+07	4.62E+07
MAM001.B3-1	2 μ L template*	2.91E+07	3.25E+07	3.08E+07
MAM001.B3-2	2 μ L template*	3.48E+07	2.92E+07	3.20E+07
MAM001.B3-3	2 μ L template*	3.57E+07	3.12E+07	3.35E+07
MAM001.B3-4	2 μ L template*	2.45E+07	2.24E+07	2.35E+07
MAM001.B3-5	2 μ L template*	3.15E+07	2.63E+07	2.89E+07
RGA001.A1	10 μ L template (standard)	2.76E+08	2.87E+08	2.82E+08
RGA001.A2	2 μ L template	4.48E+08	5.42E+08	4.95E+08
RGA001.B1	10 μ L template (standard)	2.80E+08	2.87E+08	2.84E+08
RGA001.B2	2 μ L template	3.26E+08	4.21E+08	3.74E+08

Table 1. qPCR copy numbers for pre-indexed libraries. The libraries built on only 2 μ L of extract have as many or more DNA fragments as measured in a qPCR assay after the attachment of Illumina adapters and prior to indexing and amplification. *The MAM001.A3 and MAM001.B3 libraries were constructed with 2 μ L of template per library, then combined after the indexing step.

Alpha diversity

We evaluated the species level alpha diversity of each library with species richness (count of species) and Simpson's diversity index (see METHODS). We assessed these metrics for each individual sequenced library, and for computationally combined libraries

representative of each site. For the combined libraries, we calculated diversity including and excluding human reads (see METHODS). We used MALT (Vågene et al., 2018), as implemented in the HOPS pipeline (Huebler et al., 2019), to generate metagenomic profiles for each library. MALT output was visually assessed with MEGAN6 (Huson et al., 2016), which we then used to produce separate taxon tables with assigned read numbers for species-level reads from all taxonomic groups, Bacteria, Archaea, Eukaryota, and Viruses. Because alpha diversity is a within-sample metric, assigned read counts were not normalized across the libraries.

The extract or quantity of extract used to construct each library does not appear to have played a role in the total number of species identified in the metagenome or value of the Simpson's diversity index (Figure 1). As the read numbers were not normalized for the alpha diversity calculations, the minor differences in species richness between the non-combined libraries may be due to differences in sequencing depth. The Simpson's index values for species of all taxonomic groups do not differ substantially between high template and low template libraries, with the RGA libraries appearing slightly less diverse than the MAM libraries.

We find the computationally combined libraries, while having increased species richness, do not consistently lead to increased diversity across all taxa. As Simpson's index considers evenness, and values common or more abundant species over rare species (Xia et al., 2018), we would likely not be missing the "core biome" of the sample by taking any one of the non-combined libraries. We can infer that the species unaccounted for by a single non-combined library from either site are attracting few reads, and thus are difficult to validate in downstream analyses. However, the possibility remains that by doing so, one would miss potentially important genetic evidence of a historically or archaeologically relevant organism (e.g. a pathogen or dietary item) that may be further investigated with a targeted enrichment approach (Bos et al., 2014; Schuenemann et al., 2013, 2011).

Despite the overall consistency of the Simpson's index values for all taxa, the diversity of different libraries is surprisingly dynamic for eukaryotic species. For MAM, eukaryotic diversity is stochastic across library type. For RGA, there appears to be a pattern of increased diversity for eukaryotes in the low template libraries compared to the standard template libraries. The RGA combined library with human reads removed spikes in diversity compared to its counterpart that retains human reads. This is likely due to the high quantity of *Homo sapiens* DNA present in the RGA combined library compared to other eukaryotic organisms. Of the 76,531 reads summarized to the Eukaryota node in the RGA combined library, 43,330, or approximately 57%, were assigned to *H. sapiens*. By removing all reads that map to the HG19 human reference genome, we likely increased the evenness of species representation among the eukaryotes.

The main difference between sites in terms of alpha diversity is the reduction of archaeal diversity in RGA compared to MAM libraries (Figure 1B).

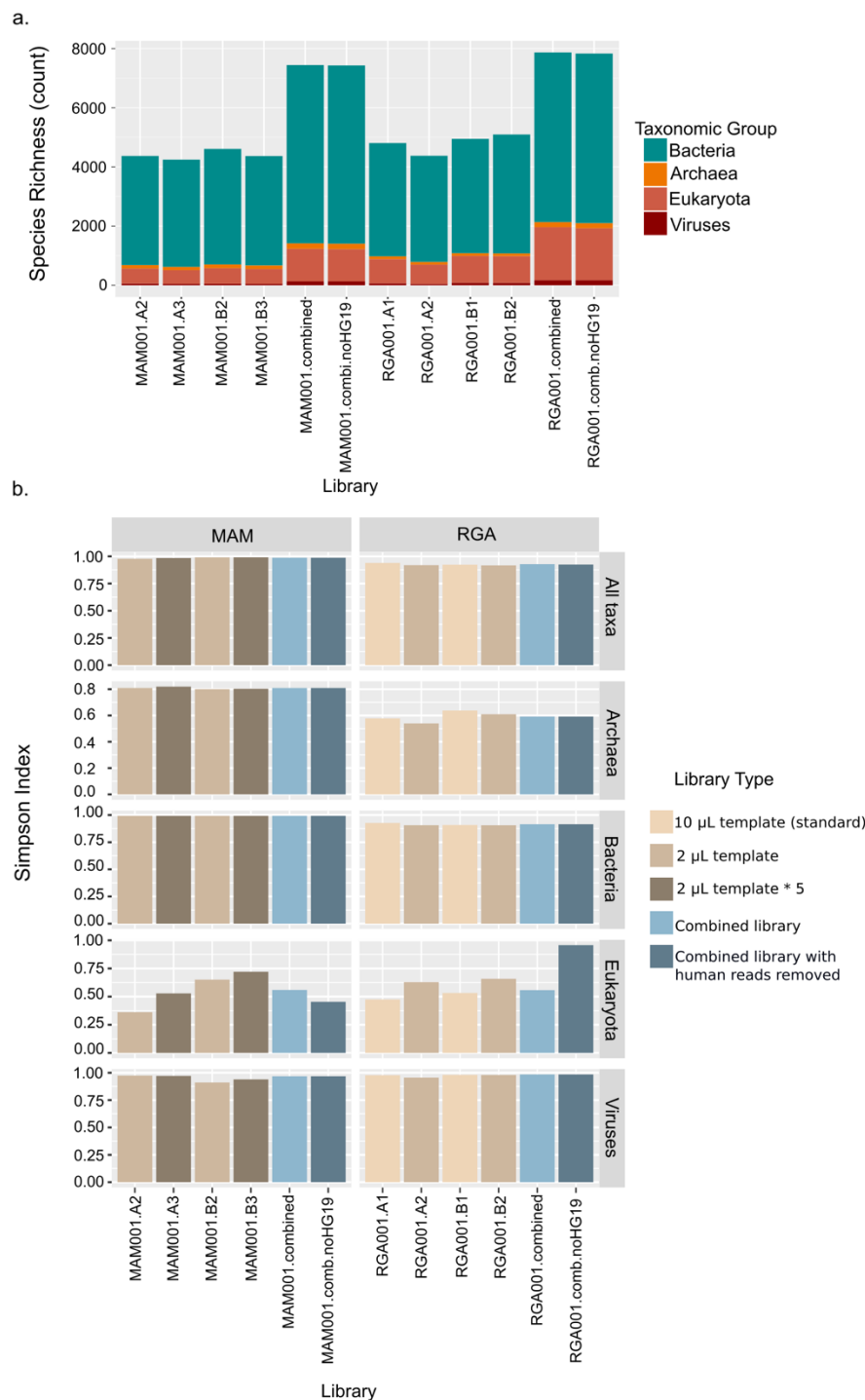


Figure 1. Alpha diversity across all sequenced libraries. (a) Species richness (count of total number of species) from metagenomic profile of each library. The bar chart is stacked by taxonomic group. **(b)** Simpson's diversity indices across all libraries for each site, divided by taxonomic group. The Simpson's index considers evenness in its formula. Indices closer to 0 have lower diversity, and indices closer to 1 have higher diversity.

Estimated source contributions to MAM and RGA bacterial and archaeal diversity

Reads summarized to the genus level were exported to a taxon table from the metagenomic profiles of the combined MAM and RGA libraries with human reads extracted using MEGAN6 (Huson et al., 2016). The taxon table was modified for compatibility with SourceTracker2 (Knights et al., 2011), a tool that produces Bayesian estimates of source contribution to a given metagenome. We modeled source contribution using reference datasets for rural human gut microbiome (Obregon-Tito et al., 2015; Rampelli et al., 2015), urban human gut microbiome (Sankaranarayanan et al., 2015), human gingival plaque (Consortium, 2012), human skin (Oh et al., 2016), and sediment (Slon et al., 2017). Human reads were also removed from the source datasets.

Both MAM and RGA were estimated to have contributions from all sources we modeled (Figure 2). However, SourceTracker2 could not estimate over 50% of contribution source to either library. This could be due to either there being a large number of taxa in MAM and RGA that are shared between multiple source models, or insufficient source models for the two sites. The model source for sediment (Slon et al., 2017) is from one location – Denisova cave – and regional differences in environmental microbiome could explain the large proportion of unknown contribution. The gut contribution estimates are almost identical, taken together. However, MAM and RGA differ in their relative proportions of urban and rural gut contribution. This could indicate differences in the gut microbiota between the two communities.

Though limited to cultured human gut bacteria, we more closely assessed the gut microbiome signature by comparing the bacterial species in our metagenomic analysis with the Human Gut Bacterial Genome and Culture Collection (HGBGCC) (Forster et al., 2019). Of the 153 bacterial species from the list present in the MAM and RGA combined libraries, 125 were present in the metagenomes from both sites (Supplementary Table 1). Among the 28 species that only appeared in one of the combined libraries, none received greater than 15 assigned reads. In all cases, it is impossible to rule out the small quantity of reads as misalignments from closely related bacteria. Of the species on the HGBGCC list, *Bifidobacterium angulatum* has the highest representation in assigned reads for both MAM and RGA. *Bifidobacterium* species constituted 6 out of the top 11 most highly represented gut bacteria species in RGA. The top species in MAM represented more diverse genera, including *Micrococcus*, *Anaerostipes*, *Ruminococcus*, and *Escherichia*.

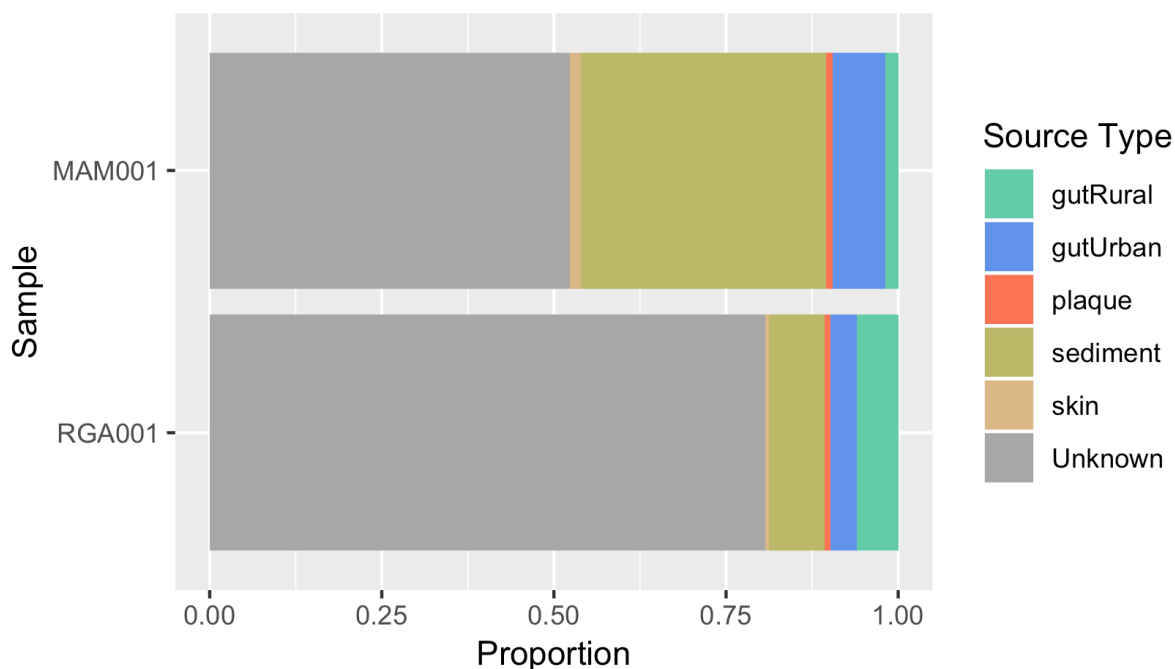


Figure 2. Estimated source contribution to MAM and RGA bacterial and archaeal

community. SourceTracker2 (Knights et al., 2011) was used to estimate contributions from five model sources: rural human gut (Obregon-Tito et al., 2015; Rampelli et al., 2015), urban human gut (Sankaranarayanan et al., 2015), human sub- and supra-gingival plaque (Consortium, 2012), human skin (Oh et al., 2016), and sediment (Slon et al., 2017). Contribution was estimated using combined libraries for each site with reads mapping to the HG19 human reference genome removed. Human reads were also removed from the model source metagenomes. Estimations are based on reads summarized to the genus level within the Bacteria and Archaea nodes.

Eukaryotic pathogens

We screened MAM and RGA for eukaryotic pathogens through the HOPS pipeline, using a customized list of target taxa. We included all taxa identified in samples from MAM and RGA previously by microscopy and ELISA assay (Yeh et al., 2015, 2014), including species within genera when identifications were made at the genus level, and close genetic relatives of the identified taxa. We also included known human-infecting parasites as listed in Ash and Orihel's *Atlas of Human Parasitology* (2015). The list was curated such that all taxon names matched those in NCBI. For each species, we included classifications on the taxonomic path up to the phylum level (e.g. for *Ascaris lumbricoides*, we included *Ascaris*, *Ascarididae*, *Ascaridoidea*, *Ascaridomorpha*, *Spirurina*, *Rhabditida*, *Chromadorea*, and *Nematoda*). Thus far, no *in silico* testing has been published to establish the specificity of simulated ancient DNA (aDNA) reads from eukaryotic pathogens as has been done for bacterial pathogens (Huebler et al., 2019).

To have confidence in the authenticity of an ancient DNA alignment, one must evaluate the sequencing read's specificity to a reference sequence and the presence of damage typical of aDNA (Huebler et al., 2019; Key et al., 2017). Though hydrolytic and oxidative damage complicate the analysis of aDNA through fragmentation and nucleotide misincorporation, cytosine-to-thymine (C>T) transitions occurring near the 5' ends of fragments (and guanine-to-adenine (G>A) transitions near the 3' ends) create a consistent pattern by which we may distinguish truly ancient DNA from modern contaminants (Briggs et al., 2007; Jesse Dabney et al., 2013; Ginolhac et al., 2011; Hofreiter et al., 2001). HOPS provides automated assessment of DNA authenticity by calculating the distribution of edit distances of reads aligning to a given reference sequence or set of reference sequences (in the case of reads being assigned to a taxonomic node above species-level) and detecting the presence and distribution of cytosine deamination along the reads. Edit distance refers to the number of mismatches between the aligned read and reference sequence. The results of these assessments are communicated through three levels of criteria: 1) edit distance distribution of all reads, 2) presence of C>T damage, and 3) edit distance distribution of all reads with damage.

Yeh and colleagues identified the following taxa by microscopy or ELISA assay in sediments and coprolites from MAM: *Trichuris trichiura*, *Ascaris lumbricoides*, *Taenia* sp., *Diphyllobothrium* (or *Dibothriocephalus*) sp., *Entamoeba histolytica*, and *Giardia duodenalis* (2015). Of these, we found evidence of authentic aDNA for *Trichuris trichiura* and *Ascaris lumbricoides*. Though *T. trichiura* passed the second threshold, only 21 reads were assigned to the node, and nucleotide misincorporation lesions consistent with damage were present only on 2 reads. The genus node of *Trichuris* has 37 assigned reads and also passed the second level of criteria. There may be authentic *T. trichiura* DNA in the sample, and we would expect to find it as it had been microscopically identified in disaggregated sediment from MAM, but at such low numbers and spread across multiple taxonomic levels, it is difficult to fully evaluate. *A. lumbricoides* passed the third level of criteria according to HOPS at the species and genus levels, and several higher taxonomic levels had assigned reads which passed level two or level one criteria (e.g. Chomadoreia, Nematoda; Figure 3). This node accumulated sufficient assigned reads (n=9389) to present a clear damage pattern. Due to this convincing evidence for the presence of *A. lumbricoides*, we mapped the metagenomic data of the combined library separately to a single reference genome (International Helminth Genomes Consortium, 2019). Approximately 0.26% of the MAM combined library aligned to the *A. lumbricoides* sequence (Figure 3b).

From the combined library for RGA, three species-level nodes passed the level two criteria: *Enterobius vermicularis*, *Dibothriocephalus latus* (also known as *Diphyllobothrium latum*), and *A. lumbricoides*. *Ascaris* and *Diphyllobothrium* had been identified previously to

the genus level by microscopy in the RGA sediments (Yeh et al., 2014). We proceeded to map the combined RGA libraries to single references for both *A. lumbricoides* and *D. latus*. Approximately 0.09% of reads from the combined RGA library aligned to *A. lumbricoides* and ~0.08% of reads aligned to *D. latus* (Figures 3c and 3d, Table 2). Of the taxa identified by Yeh and colleagues, there was no genetic evidence for *Oxyuris equi*. Reads aligning to *T. trichiura* passed the first threshold. *E. vermicularis*, or human pinworm, had not been identified morphologically in prior investigations. However, it passed the second threshold with only 17 reads, limiting further authentication.

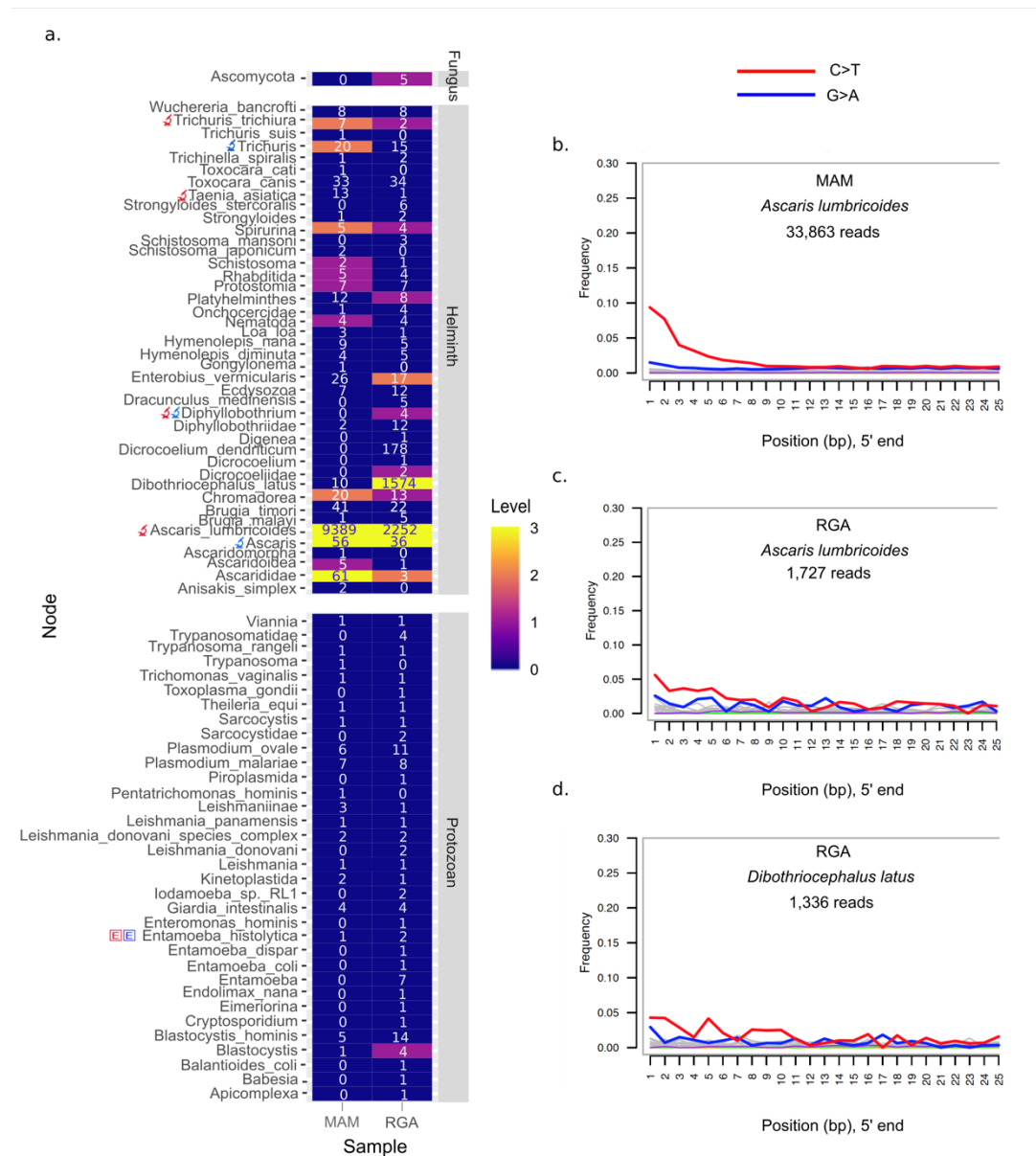


Figure 3. Eukaryotic parasite taxa with read alignments from latrine samples. (a) The nodes labelled to the left of the heatmap correspond to taxa that are or contain eukaryotic parasitic species. All taxon nodes listed here have at least one assigned read aligning to associated sequences from the MALT results of the MAM and RGA combined libraries (with reads aligning to the HG19 human reference genome removed). All taxa in this figure were

assigned reads in at least one of the combined libraries, and the assigned read count (not normalized) can be found in the respective cell in the heatmap. The level indicated by the color of each cell refers to how many authentication criteria are reached by the alignment as determined with HOPS. Level 0 – no criteria were passed, level 1 – edit distance criterion was passed, level 2 – damage criterion was passed, level 3 – edit distance of damaged reads criterion was passed. The blue and red microscope symbols indicate the taxon was identified by microscopy in samples from RGA or MAM, respectively. The blue and red “E” indicate the taxon was identified by ELISA assay. **(b-d)** Damage plots of combined library data mapped to *Ascaris lumbricoides* and *Dibothriocephalus latus*. Reads indicate reads mapped to the reference sequence after quality filtering (see METHODS).

	Sample	Total Reads	Mapped Reads	Endogenous DNA (%)	Cluster Factor	Mean Coverage
<i>A. lumbricoides</i>	MAM	16463260	33863	0.264	1.068	0.0087
<i>A. lumbricoides</i>	RGA	21202878	1727	0.090	1.101	0.0004
<i>D. latus</i>	RGA	21202878	1336	0.082	1.113	0.0002

Table 2. Mapping statistics for *Ascaris lumbricoides* and *Dibothriocephalus latus*.

Mapping statistics were reported in the EAGER report table (Peltzer et al., 2016). Mapping was performed in EAGER with BWA (see METHODS). Mapped reads indicate the number of reads mapped to the reference sequence following quality filtering and the removal of duplicates.

DISCUSSION

Template volume did not impact metagenomic alpha diversity

Despite the inhibition phenomenon we detected during our quantitation of copy number in the pre-indexed libraries (Table 1), we did not observe an impact on the alpha diversity measurements of the resulting metagenomes. We can infer that the standard library preparation method for the RGA samples did not result in a meaningful decrease in diversity (as calculated with the Simpson’s index) for all taxa together, Bacteria, Archaea, or Viruses. However, we do see an increase in Eukaryotic diversity for the 2 μ L libraries over the 10 μ L libraries. The mechanism behind this signature is unclear. No discernable pattern emerged regarding the relationship between template volume and alpha diversity for the 2 μ L and 2 μ L*5 libraries. It is unclear whether this lack of relationship stems from inhibition or sufficiency of the low template volume to produce a representative metagenome for the sample. However, the relatively low cluster factors we see in the mapping statistics for *A. lumbricoides* and *D. latus*, a measure of how many times one is likely to have seen a given read based on the duplication rate of mapped reads, indicate the combined sequencing depth for each site has not exhausted the complexity of the libraries (Table 2).

Gut microbiome preservation and prospects for regional comparisons

The SourceTracker2 results presented here indicate a) a large proportion of bacterial and archaeal DNA in MAM and RGA does not share an affinity with any of the source models we used, and b) MAM and RGA contain approximately the same proportion of DNA from the human gut, but share different proportions of the estimated human gut input between “rural” and “urban” microbial profiles (Figure 2). The unknown contribution to MAM and RGA could be due to our use of a single dataset from Denisova cave as the environmental “soil” source (Slon et al., 2017). This may not be representative of bacterial and archaeal diversity in the local soil at either site. The difference in estimated rural vs. urban gut contribution is intriguing given the similarity in total gut contribution between MAM and RGA. Based on our cross-referencing of taxa present in the libraries with the HGBGCC list, there is no major bacterial gut species represented at one site but not the other. The SourceTracker2 analysis was performed on genus-level summarized reads, so the difference between MAM and RGA may stem from different levels of representation of certain genera. For instance, *Bifidobacterium* species are more dominant in RGA than in MAM.

Recent efforts to characterize the unculturable component of the human gut microbiome will likely soon lead to the expansion of databases to include a broader range of taxa (Almeida et al., 2019; Pasolli et al., 2019). This could lead to better understanding not only of what bacteria and archaea tend to naturally inhabit the human gut, but also of the importance of community structure, and how it correlates with behavior (e.g. diet or lifestyle), life stage, geography, or ancestry. In addition, further human microbiome studies among non-industrialized populations, which we know tend to have different gut microbiome makeup than populations from industrialized societies (Filippo et al., 2010; Obregon-Tito et al., 2015), will offer datasets that would work as better references for analyzing ancient pre-industrialization populations in the archaeological record. We could learn more from the existing MAM and RGA shotgun sequences as greater understanding and characterization of the human gut microbiome becomes available.

Eukaryotic parasites: the promise and limitations of ancient metagenomics

Of the eukaryotic parasites previously morphologically identified in sediment and coprolites from MAM and RGA, we made confident identifications of *Ascaris lumbricoides* in MAM and *A. lumbricoides* and *Dibothriocephalus latus* in RGA. A small quantity of reads from each library was assigned to the other previously identified eukaryotic parasites (or closely-related taxa) with the exception of *Oxyuris equi*, though these were not sufficient to allow authentication (Figure 3). *Enterobius vermicularis* received assigned reads from MAM and RGA, and the 17 reads from RGA passed the second threshold of authentication criteria determined in HOPS, meaning the assigned reads sufficiently fit the preferred edit distance distribution from the reference sequence, and some reads contained C>T damage we would

expect to find in authentic aDNA (Huebler et al., 2019). This is the only organism that was not previously identified for which assigned reads passed the second criteria. Its presence would be expected in metagenomic datasets including human fecal material as *E. vermicularis* and other species with *Enterobius* maintain reservoirs in humans and non-human primates (Araújo et al., 2013; Mitchell, 2013).

The low quantity of DNA from the previously identified eukaryotic parasites is likely due in part to our approach to sample preparation and our non-targeted metagenomic analysis. Prior genetic investigations into ancient eukaryotic parasites (largely dominated by *A. lumbricoides*) have generally been targeted either through the direct extraction and sequencing of identified and filtered parasite eggs, or through amplicon sequencing of specific genes belonging to a single species or small selection of taxa (Côté et al., 2016; Guhl et al., 1999; Iñiguez et al., 2006, 2003; Maicher et al., 2017; Oh et al., 2010b, 2010a; Søre et al., 2018). Here, we aimed to avoid contamination of the sample by extracting aliquots of the bulk sample directly, foregoing filtration of the sediments for eggs, and not limiting our analyses by selective amplification. Given the extremely low coverage for even the most strongly represented eukaryotic pathogens in this dataset, what is the benefit of a metagenomics approach to paleoparasitology?

Though morphology is a valuable and reliable tool for establishing the presence of parasites that leave eggs of sufficient size in the archaeological record, fine-grain taxonomic identification is not possible for some groups (e.g. *Entamoeba* spp.) and taphonomic processes may complicate identification (Côté et al., 2016; Mitchell, 2013). Genetic data may offer such fine-grain taxonomic identification. In this case, a metagenomic sequencing coupled with a lowest common ancestor taxonomic binning approach (Søre et al., 2018), as opposed to a marker gene or amplicon approach (Anderson, 2001), may offer more robust species identifications. Following relatively unbiased identification of an organism, coverage may be increased through deeper sequencing or a genomic enrichment strategy may be designed. The resulting data could be relevant for elucidating the evolutionary history of a parasite. However, the full promise of paleoparasitology by metagenomics to answer high resolution questions of parasite phylogeny and evolution cannot be fulfilled until more complete and reliable modern reference databases are available. For instance, differentiating between *Dibothriocephalus latus* and *Diphyllbothrium dendriticum*, two closely related but distinct species of fish tapeworm, with low coverage metagenomic data would likely bias for *Dibothriocephalus latus* due to the increased volume of full genome data available. Groups such as the International Helminth Genomes Consortium are working to fill existing database gaps (International Helminth Genomes Consortium, 2019).

MATERIALS AND METHODS

Sampling and DNA extraction

Un-sieved portions of bulk sediment (~3 g per site) from medieval latrines in Riga and Jerusalem were sent from the University of Cambridge, where different portions of the sediment had been microscopically and chemically analyzed (Yeh et al., 2015, 2014), to the Max Planck Institute for Science of Human History (MPI-SHH) in Jena, Germany. Further sample preparation, DNA extraction, Illumina library preparation, and sequencing were conducted at the dedicated ancient DNA clean room (pre-PCR) and post-PCR facilities of the MPI-SHH. Two spatially separate portions of approximately 50-80 mg of sediment were taken from each sample, and a total of four extractions were performed. Extractions were performed using a protocol optimized for use with archaeological bone (Jesse Dabney et al., 2013), adapted for sediment after (Slon et al., 2017). Each sub-sample was immersed in 1 mL of 0.5 M EDTA and rotated overnight at room temperature. After decalcification, the sample was centrifuged and the supernatant was purified using a High Pure Viral Nucleic Acid Large Volume kit (Roche) and a 5 M guanidine-hydrochloride binding buffer. Elution was performed using 100 µl of a 10 mM tris-hydrochloride, 1 mM EDTA (pH 8.0), and 0.05% Tween-20 buffer (TET). The eluate from the MAM extracts had a dark-brown tinge, and that of the RGA extracts had a very slight pale yellow-tinge, indicating impurities in the extract. Two negative controls were included in the extraction batch to detect reagent- and process-derived contamination.

Library preparation

Initial screening and inhibition test

Construction of double-stranded Illumina libraries was performed according to a modified established protocol (Meyer and Kircher, 2010). DNA overhangs were removed and filled-in with a 50 µl reaction including 5 µl of NEB Buffer 2 (New England Biolabs), 2 µl dNTP mix (2.5 mM), 4 µl BSA (10 mg/mL), 5 µl ATP (10 mM), 2 µl T4 polynucleotide kinase, and 0.4 µl T4 polymerase. Four libraries (MAM001.A1, MAM001.B1, RGA001.A1, and RGA001.B1) were constructed using 10 µl of DNA extract and 21.6 µl H₂O. Two additional libraries (MAM001.A2 and MAM001.B2) were constructed using 2 µl of DNA extract and 29.6 µl H₂O in response to the coloring of the MAM extracts. We wished to test if chemicals remaining in the MAM extracts would inhibit downstream PCR steps by lowering the template volume and comparing copy number calculations after quantitating the non-indexed libraries with a real-time qPCR assay (Lightcycler 480 Roche; see below). Following the incubation of the initial overhang repair reactions, they were purified over MinElute columns (Qiagen) and eluted in 18 µl TET. Universal Illumina adapters were ligated to the repaired fragments with a reaction of 20 µl Quick Ligase Buffer, 1 µl Illumina adapter mix (0.25 µM), and 1 µl of Quick Ligase. The reaction was then purified as described above and eluted in 20 µl of TET. Adapter fill-in was performed in 40 µl reactions, including 4 µl Thermopol buffer, 2 µl dNTP

mix (2.5 mM) and 2 µl Bst polymerase. Following incubation at 37°C for 20 minutes, the enzyme was heat-deactivated with an additional 20-minute incubation at 80°C. Two negative controls were included to the workflow to track contamination specifically in the library preparation reagents and process. The resulting pre-indexed libraries were quantitated with RT-qPCR, at which point inhibition in the MAM libraries was evaluated (see RESULTS). The MAM libraries built with 2 µl of template (MAM001.A2 and MAM001.B2) were carried forward for further processing.

Later, five pre-indexed libraries were constructed from each extract for MAM using 2 µl of template each (MAM001.A3-1, MAM001.A3-2, MAM001.A3-3, MAM001.A3-4, MAM001.A3-5, MAM001.B3-1, MAM001.B3-2, MAM001.B3-3, MAM001.B3-4, MAM001.B3-5), and one pre-indexed library were constructed from each extract for RGA using 2 µl of template to test whether we had an inhibition effect in the qPCR results for RGA as well.

Indexing, amplification, and sequencing

All pre-indexed libraries that were carried forward for sequencing were double indexed with unique pairs of indices over 100 µl reactions containing less than 1.5E+10 DNA fragments. All libraries referred to as MAM001.A3 were given the same indices. All libraries referred to as MAM001.B3 were given the same indices. Each reaction included 10 µl PfuTurbo buffer, 1 µl PfuTurbo (Agilent), 1 µl dNTP mix (25mM), 1.5 µl BSA (10 mg/ml), and 2 µl of each indexing primer (10 µM). The master mix was prepared in a pre-PCR clean room and transported to a separate lab for amplification. All reactions belonging to the same library were purified over MinElute columns. The indexed MAM001.A3 sub-libraries and MAM001.B3 sub-libraries were combined during purification. Indexed libraries were eluted in 50 µl of TET. Indexing efficiency was assessed using a qPCR assay targeting the IS5 and IS6 sequences in the indexing primers. Approximately one-third of each indexed library was amplified with Herculase II Fusion DNA Polymerase (Agilent). The products were MinElute purified and quantified using an Agilent Tape Station D1000 Screen Tape kit. MAM001.A2, MAM001.B2, RGA001.A1, and RGA001.A2 were sequenced on an Illumina NextSeq 500 using a paired-end, 75-cycle high-output kit. MAM001.A3, MAM001.B3, RGA001.A2, and RGA001.B2 were sequenced on an Illumina NextSeq 500 using a single-end, 75-cycle high-output kit. Negative controls were sequenced separately to avoid index cross-talk.

Computational methods

Pre-processing and mapping to HG19

We pre-processed and mapped de-multiplexed reads for each library using the EAGER pipeline (v (Peltzer et al., 2016)). Preliminary adapter-removal and read filtering was performed using AdapterRemoval, enforcing a minimum read length of 30 and a minimum base quality of 20. Paired reads for MAM001.A0102, MAM001.B0102, RGA001.A0101, and RGA001.B0101 were also merged. All reads were mapped to the HG19 human reference

genome using BWA (Li and Durbin, 2009) (as implemented in EAGER) with a quality filter of 30, difference parameter (-n) of 0.01, and deactivated seed length filter (-l). We performed this mapping as a standard exploratory analysis of human DNA preservation. In addition, we concatenated the fastq files from libraries belonging to the same site and mapped these combined files (MAM001.combined and RGA001.combined) to the HG19 reference genome using BWA outside of the EAGER pipeline in order to generate a fastq file with all human reads extracted. We did this to test the impact of removing human reads on the metagenomic profile of the latrines, as we expect some genetic sequences from human-associated parasites to be contaminated with human DNA. For this analysis, we applied -n 0.1 and -l 32.

HOPS

We executed a broad screening of the total genetic content of the shotgun-sequenced libraries using the HOPS pipeline (Huebler et al., 2019) which joins the taxonomic binning tool MALT (Vågene et al., 2018) with MaltExtract and a post-processing script which produces a heat-map of putatively authentic taxon identifications based on edit distance from a reference sequence and the presence of damaged reads. As input, we used fastq files from the pre-processing step above, including the original fastq files representing individual libraries, the combined fastq files, and the combined fastq files with human-mapped reads extracted. For MALT (v 040), we used the NCBI full nucleotide database ('nt', October 2017) with a 90% identity threshold. We curated a taxonomy list designed to target a) bacterial and viral pathogens of broad interest, b) human-associated eukaryotic parasites (Ash and Orihel, 2015), and c) parasites previously identified by microscopy in the Jerusalem and Riga latrine deposits (Yeh et al., 2015, 2014). The taxon names for the eukaryotic pathogens were checked against the NCBI database to ensure nomenclature matched between the taxon list and the database (Supplementary Table 1). Taxonomic levels up to phylum were included in the list for Eukaryotic pathogens (e.g. for the target species *Ascaris lumbricoides*, we included *Ascaris*, *Ascarididae*, *Ascaridoidea*, *Ascaridomorpha*, *Spirurina*, *Rhabditida*, *Chromadorea*, *Nematoda*, *Ecdysozoa*, and *Protostomia*).

Mapping to *Ascaris lumbricoides* and *Dibothriocephalus latus*

Following HOPS analysis, we mapped the combined libraries for MAM and RGA to genomic assemblies for *Ascaris lumbricoides* and *Dibothriocephalus latus* with BWA (Li and Durbin, 2009) as implemented in the EAGER pipeline (v 1.92) (Peltzer et al., 2016). As the reference for *A. lumbricoides*, we used the genome available from WormBase as of 13 February 2019 ("Ascaris lumbricoides - WormBase ParaSite," n.d.; International Helminth Genomes Consortium, 2019). As the reference for *Dibothriocephalus latus* we used the full genomic assembly available under accession number PRJEB1206 (Park et al., 2007). For both organisms we used the following mapping parameters: -l 16, -n 0.01, and -q 37.

Duplicates were removed with MarkDuplicates, and damage profiling was performed with mapDamage (Ginolhac et al., 2011).

Metagenomic Profiling

We generated interactive donut plots to explore the total metagenomic profile for the combined libraries (excluding human reads) for MAM and RGA using Krona (Ondov et al., 2011). For Figures 1 and 2, we limited visualization to a depth of four taxonomic levels to ensure readability. Additionally, we used SourceTracker2 to estimate source contributions to the sites' metagenomes utilizing five model sources: rural human gut, urban human gut, human sub- and supra-gingival plaque, human skin, and sediment (Slon et al., 2017). We used MALT results (see specifications above) from human read-filtered fastq files of the model sources and the combined MAM and RGA libraries. The quantity of reads summarized to the genus level within the Bacteria and Archaea nodes were extracted into a taxon table using MEGAN6 (Huson et al., 2016). The table was edited to be compatible with SourceTracker2.

Community diversity

We calculated species richness and Simpson's index of diversity for each library with species-level taxon tables extracted using MEGAN6 (Huson et al., 2016). Alpha diversity calculations were performed with the *vegan* package in R (Oksanen et al., 2019). The following taxon tables were extracted: all terminal leaves (in all domains of life) at species level, leaves below the bacteria and archaea nodes at species level, leaves below the Eukaryota node at species level, leaves below the Viruses node at species level, all leaves at genus level, leaves below the cellular organisms node at genus level, leaves below the Bacteria and Archaea nodes at genus level, leaves below the Eukaryota node at genus level, and leaves below the Viruses node at genus level. Here, we define Simpson's index of diversity as

$$D = 1 - \sum_{i=1}^S p_i^2$$

where P_i is the proportion of individuals of species i in the community (He and Hu, 2005).

ACKNOWLEDGMENTS

James Fellows Yates assisted with data analysis. Guido Brandt and Antje Wissgott facilitated sequencing of the Illumina libraries and manage the laboratory facilities at the Max Planck Institute for the Science of Human History (MPI-SHH). We would like to thank the Molecular Paleopathology and Computational Pathogenomics groups at the MPI-SHH for support and discussion.

AUTHOR CONTRIBUTIONS

Study was conceived by P.M. and K.I.B. H.-Y.Y. performed microscopy and ELISA analysis of sediments from MAM and RGA. S.S. performed genetics laboratory work and metagenomic analysis. S.S. composed manuscript.

COMPETING INTERESTS

We declare no competing interests.

REFERENCES

- Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., Finn, R.D., 2019. A new genomic blueprint of the human gut microbiota. *Nature* 568, 499. <https://doi.org/10.1038/s41586-019-0965-1>
- Anderson, T.J.C., 2001. The dangers of using single locus markers in parasite epidemiology: *Ascaris* as a case study. *Trends in Parasitology* 17, 183–188. [https://doi.org/10.1016/S1471-4922\(00\)01944-9](https://doi.org/10.1016/S1471-4922(00)01944-9)
- Araújo, A., Ferreira, L.F., 2000. Paleoparasitology and the antiquity of human host-parasite relationships. *Mem. Inst. Oswaldo Cruz* 95, 89–93. <https://doi.org/10.1590/S0074-02762000000700016>
- Araujo, A., Reinhard, K., Ferreira, L.F., Pucu, E., Chieffi, P.P., 2013. Paleoparasitology: the origin of human parasites. *Arq. Neuro-Psiquiatr.* 71, 722–726. <https://doi.org/10.1590/0004-282X20130159>
- Araujo, A., Reinhard, K.J., Ferreira, L.F., Gardner, S.L., 2008. Parasites as probes for prehistoric human migrations? *Trends in Parasitology* 24, 112–115. <https://doi.org/10.1016/j.pt.2007.11.007>
- Ascaris lumbricoides* - WormBase ParaSite [WWW Document], n.d. URL https://parasite.wormbase.org/Ascaris_lumbricoides_prjeb4950/Info/Index (accessed 5.20.19).
- Ash, L.R., Orihel, T.C., 2015. *Ash & Orihel's Atlas of Human Parasitology*, 5th ed. American Society for Clinical Pathology Press.
- Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514, 494–497. <https://doi.org/10.1038/nature13591>
- Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *PNAS* 104, 14616–14621. <https://doi.org/10.1073/pnas.0704665104>
- Consortium, T.H.M.P., 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. <https://doi.org/10.1038/nature11234>
- Côté, N.M.L., Daligault, J., Pruvost, M., Bennett, E.A., Gorgé, O., Guimaraes, S., Capelli, N., Bailly, M.L., Geigl, E.-M., Grange, T., 2016. A New High-Throughput Approach to Genotype Ancient Human Gastrointestinal Parasites. *PLOS ONE* 11, e0146230. <https://doi.org/10.1371/journal.pone.0146230>
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 110, 15758–15763. <https://doi.org/10.1073/pnas.1314445110>
- Dabney, Jesse, Meyer, M., Pääbo, S., 2013. Ancient DNA Damage. *Cold Spring Harb Perspect Biol* 5, a012567. <https://doi.org/10.1101/cshperspect.a012567>
- Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by

- a comparative study in children from Europe and rural Africa. *PNAS* 107, 14691–14696. <https://doi.org/10.1073/pnas.1005963107>
- Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire, T.T., Zhu, A., Shao, Y., Pike, L.J., Louie, T., Browne, H.P., Mitchell, A.L., Neville, B.A., Finn, R.D., Lawley, T.D., 2019. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nature Biotechnology* 37, 186. <https://doi.org/10.1038/s41587-018-0009-7>
- Fugassa, M.H., Beltrame, M.O., Sardella, N.H., Civalero, M.T., Aschero, C., 2010. Paleoparasitological results from coprolites dated at the Pleistocene–Holocene transition as source of paleoecological evidence in Patagonia. *Journal of Archaeological Science* 37, 880–884. <https://doi.org/10.1016/j.jas.2009.11.018>
- Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E., Orlando, L., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27, 2153–2155. <https://doi.org/10.1093/bioinformatics/btr347>
- Gonçalves, M.L.C., Araújo, A., Duarte, R., da Silva, J.P., Reinhard, K., Bouchet, F., Ferreira, L.F., 2002. Detection of *Giardia duodenalis* antigen in coprolites using a commercially available enzyme-linked immunosorbent assay. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 96, 640–643. [https://doi.org/10.1016/S0035-9203\(02\)90337-8](https://doi.org/10.1016/S0035-9203(02)90337-8)
- Guhl, F., Jaramillo, C., Vallejo, G.A., Yockteng, R., CARDENAS-ARROYO, F., Fornaciari, G., Arriaza, B., Aufderheide, A.C., 1999. Isolation of *Trypanosoma cruzi* DNA in 4, 000-year-old mummified human tissue from northern Chile. *American Journal of Physical Anthropology* 108, 401–407.
- Han, E.-T., Guk, S.-M., Kim, J.-L., Jeong, H.-J., Kim, S.-N., Chai, J.-Y., 2003. Detection of parasite eggs from archaeological excavations in the Republic of Korea. *Mem. Inst. Oswaldo Cruz* 98, 123–126. <https://doi.org/10.1590/S0074-02762003000900018>
- He, F., Hu, X.-S., 2005. Hubbell's fundamental biodiversity parameter and the Simpson diversity index. *Ecology Letters* 8, 386–390. <https://doi.org/10.1111/j.1461-0248.2005.00729.x>
- Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. von, Pääbo, S., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29, 4793–4799. <https://doi.org/10.1093/nar/29.23.4793>
- Huebler, R., Key, F.M., Warinner, C., Bos, K.I., Krause, J., Herbig, A., 2019. HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *bioRxiv* 534198. <https://doi.org/10.1101/534198>
- Huson, D.H., Beier, S., Flade, I., Górská, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* 12, e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>
- Iñiguez, A.M., Reinhard, K., Carvalho Gonçalves, M.L., Ferreira, L.F., Araújo, A., Paulo Vicente, A.C., 2006. SL1 RNA gene recovery from *Enterobius vermicularis* ancient DNA in pre-Columbian human coprolites. *International Journal for Parasitology* 36, 1419–1425. <https://doi.org/10.1016/j.ijpara.2006.07.005>
- Iñiguez, A.M., Reinhard, K.J., Araújo, A., Ferreira, L.F., Vicente, A.C.P., 2003. *Enterobius vermicularis*: ancient DNA from north and south American human coprolites. *Memórias do Instituto Oswaldo Cruz* 98, 67–69. <https://doi.org/10.1590/S0074-02762003000900013>
- International Helminth Genomes Consortium, 2019. Comparative genomics of the major parasitic worms. *Nature Genetics* 51, 163. <https://doi.org/10.1038/s41588-018-0262-1>
- Key, F.M., Posth, C., Krause, J., Herbig, A., Bos, K.I., 2017. Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. *Trends in Genetics* 33, 508–520. <https://doi.org/10.1016/j.tig.2017.05.005>
- Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-

- independent microbial source tracking. *Nat Meth* 8, 761–763. <https://doi.org/10.1038/nmeth.1650>
- Le Bailly, M., Gonçalves, M.L., Harter-Lailheugue, S., Prodéo, F., Araujo, A., Bouchet, F., 2008. New finding of *Giardia intestinalis* (Eukaryote, Metamonad) in Old World archaeological site using immunofluorescence and enzyme-linked immunosorbent assays. *Mem. Inst. Oswaldo Cruz* 103, 298–300. <https://doi.org/10.1590/S0074-02762008005000018>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Loreille, O., Roumat, E., Verneau, O., Bouchet, F., Hänni, C., 2001. Ancient DNA from *Ascaris*: extraction amplification and sequences from eggs collected in coprolites. *International Journal for Parasitology* 31, 1101–1106. [https://doi.org/10.1016/S0020-7519\(01\)00214-4](https://doi.org/10.1016/S0020-7519(01)00214-4)
- Maicher, C., Hoffmann, A., Côté, N.M., Palomo Pérez, A., Saña Seguí, M., Le Bailly, M., 2017. Paleoparasitological investigations on the Neolithic lakeside settlement of La Draga (Lake Banyoles, Spain). *The Holocene* 0959683617702236.
- Meyer, M., Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* 2010, pdb.prot5448-pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Mitchell, P.D., 2013. The origins of human parasites: Exploring the evidence for endoparasitism throughout human evolution. *International Journal of Paleopathology* 3, 191–198. <https://doi.org/10.1016/j.ijpp.2013.08.003>
- Mitchell, P.D., Anastasiou, E., Syon, D., 2011. Human intestinal parasites in crusader Acre: Evidence for migration with disease in the medieval period. *International Journal of Paleopathology* 1, 132–137. <https://doi.org/10.1016/j.ijpp.2011.10.005>
- Mitchell, P.D., Stern, E., Tepper, Y., 2008. Dysentery in the crusader kingdom of Jerusalem: an ELISA analysis of two medieval latrines in the City of Acre (Israel). *Journal of Archaeological Science* 35, 1849–1853. <https://doi.org/10.1016/j.jas.2007.11.017>
- Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M., Spicer, P., Lawson, P., Marin-Reyes, L., Trujillo-Villarreal, O., Foster, M., Gujja-Poma, E., Troncoso-Corzo, L., Warinner, C., Ozga, A.T., Lewis, C.M., 2015. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Communications* 6, 6505. <https://doi.org/10.1038/ncomms7505>
- Oh, C.S., Seo, M., Chai, J.Y., Lee, S.J., Kim, M.J., Park, J.B., Shin, D.H., 2010a. Amplification and sequencing of *Trichuris trichiura* ancient DNA extracted from archaeological sediments. *Journal of Archaeological Science* 37, 1269–1273. <https://doi.org/10.1016/j.jas.2009.12.029>
- Oh, C.S., Seo, M., Lim, N.J., Lee, S.J., Lee, E.-J., Lee, S.D., Shin, D.H., 2010b. Paleoparasitological report on *Ascaris* aDNA from an ancient East Asian sample. *Memórias do Instituto Oswaldo Cruz* 105, 225–228. <https://doi.org/10.1590/S0074-02762010000200020>
- Oh, J., Byrd, A.L., Park, M., Kong, H.H., Segre, J.A., 2016. Temporal Stability of the Human Skin Microbiome. *Cell* 165, 854–866. <https://doi.org/10.1016/j.cell.2016.04.008>
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. *vegan: Community Ecology Package*.
- Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 385. <https://doi.org/10.1186/1471-2105-12-385>
- Park, J.-K., Kim, K.-H., Kang, S., Jeon, H.K., Kim, J.-H., Littlewood, D.T.J., Eom, K.S., 2007. Characterization of the mitochondrial genome of *Diphyllobothrium latum* (Cestoda: Pseudophyllidea) - implications for the phylogeny of eucestodes. *Parasitology* 134, 749–759. <https://doi.org/10.1017/S003118200600206X>

- Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., DuLong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., Segata, N., 2019. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 176, 649-662.e20. <https://doi.org/10.1016/j.cell.2019.01.001>
- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., Nieselt, K., 2016. EAGER: efficient ancient genome reconstruction. *Genome Biology* 17, 60. <https://doi.org/10.1186/s13059-016-0918-z>
- Perry, G.H., 2014. Parasites and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews* 23, 218–228. <https://doi.org/10.1002/evan.21427>
- Rampelli, S., Schnorr, S.L., Consolandi, C., Turrioni, S., Severgnini, M., Peano, C., Brigidi, P., Crittenden, A.N., Henry, A.G., Candela, M., 2015. Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Current Biology* 25, 1682–1693. <https://doi.org/10.1016/j.cub.2015.04.055>
- Reinhard, K.J., Hevly, R.H., Anderson, G.A., 1987. Helminth Remains from Prehistoric Indian Coprolites on the Colorado Plateau. *The Journal of Parasitology* 73, 630. <https://doi.org/10.2307/3282147>
- Sankaranarayanan, K., Ozga, A.T., Warinner, C., Tito, R.Y., Obregon-Tito, A.J., Xu, J., Gaffney, P.M., Jervis, L.L., Cox, D., Stephens, L., Foster, M., Tallbull, G., Spicer, P., Lewis, C.M., 2015. Gut Microbiome Diversity among Cheyenne and Arapaho Individuals from Western Oklahoma. *Current Biology* 25, 3161–3169. <https://doi.org/10.1016/j.cub.2015.10.060>
- Schuenemann, V.J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mitnik, A., Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J.D., White, W., Krause, J., Poinar, H.N., 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proceedings of the National Academy of Sciences* 108, E746–E752. <https://doi.org/10.1073/pnas.1105107108>
- Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J.L., Kjellström, A., Wu, H., Stewart, G.R., Taylor, G.M., Bauer, P., Lee, O.Y.-C., Wu, H.H.T., Minnikin, D.E., Besra, G.S., Tucker, K., Roffey, S., Sow, S.O., Cole, S.T., Nieselt, K., Krause, J., 2013. Genome-Wide Comparison of Medieval and Modern *Mycobacterium leprae*. *Science* 341, 179–183. <https://doi.org/10.1126/science.1238286>
- Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., Rasilla, M. de la, Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., Stewart, J.R., Derevianko, A.P., Jacobs, Z., Li, B., Roberts, R.G., Shunkov, M.V., Lumley, H. de, Perrenoud, C., Gušić, I., Kučan, Ž., Rudan, P., Aximu-Petri, A., Essel, E., Nagel, S., Nickel, B., Schmidt, A., Prüfer, K., Kelso, J., Burbano, H.A., Pääbo, S., Meyer, M., 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* eaam9695. <https://doi.org/10.1126/science.aam9695>
- Søe, M.J., Nejsum, P., Seersholm, F.V., Fredensborg, B.L., Habraken, R., Haase, K., Hald, M.M., Simonsen, R., Højlund, F., Blanke, L., Merkyte, I., Willerslev, E., Kapel, C.M.O., 2018. Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. *PLOS ONE* 13, e0195481. <https://doi.org/10.1371/journal.pone.0195481>
- Vågene, Å.J., Herbig, A., Campana, M.G., García, N.M.R., Warinner, C., Sabin, S., Spyrou, M.A., Valtueña, A.A., Huson, D., Tuross, N., Bos, K.I., Krause, J., 2018. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution* 520–528. <https://doi.org/10.1038/s41559-017-0446-6>
- Xia, Y., Sun, J., Chen, D.-G., 2018. *Statistical analysis of microbiome data with R*. Springer Berlin Heidelberg, New York, NY.
- Yeh, H.-Y., Pluskowski, A., Kalējs, U., Mitchell, P.D., 2014. Intestinal parasites in a mid-14th century latrine from Riga, Latvia: fish tapeworm and the consumption of uncooked fish in the medieval eastern Baltic region. *Journal of Archaeological Science* 49, 83–89. <https://doi.org/10.1016/j.jas.2014.05.001>

Yeh, H.-Y., Prag, K., Clamer, C., Humbert, J.-B., Mitchell, P.D., 2015. Human intestinal parasites from a Mamluk Period cesspool in the Christian quarter of Jerusalem: Potential indicators of long distance travel in the 15th century AD. *International Journal of Paleopathology* 9, 69–75. <https://doi.org/10.1016/j.ijpp.2015.02.003>

SUPPLEMENTARY TABLE 1

Taxon	Level	NCBI_TaxonID
Acanthamoeba	genus	5754
Acanthocephala	phylum	10232
Aconoidasida	class	422676
Anatrichosoma	genus	119100
Ancylostoma	genus	29169
Ancylostoma duodenale	species	51022
Anisakidae	family	6267
Anisakis simplex	species	6269
Apansporoblastina	suborder	6032
Apicomplexa	phylum	5794
Archiacanthocephala	class	10233
Ascarididae	family	6250
Ascaridoidea	superfamily	33256
Ascaridomorpha	infraorder	6249
Ascaris	genus	6251
Ascaris lumbricoides	species	6252
Ascomycota	phylum	4890
Babesia	genus	5864
Babesia divergens	species	32595
Babesia microti	species	5868
Babesiidae	family	32594
Balamuthia	genus	66526
Balamuthia mandrillaris	species	66527
Balamuthiidae	family	555408
Balantioides	genus	2038102
Balantidium coli	species	71585
Baylisascaris procyonis	species	6259
Blastocystis	genus	12967
Blastocystis hominis	species	12968
Brugia malayi	species	6279
Brugia timori	species	42155
Capillaria	genus	119095
Capillaria hepatica	species	1239592
Centramoebida	noRank	555407
Cestoda	class	6199
Chabertiidae	family	52651
Chilomastix mesnili	species	472572
Chromadorea	class	119089
Cloacinidae	family	63235
Clonorchis	genus	79922

Clonorchis sinensis	species	79923
Coccidia	subclass	5796
Conoidasida	class	1280412
Cryptosporidiidae	family	35082
Cryptosporidium	genus	5806
Cryptosporidium hominis	species	237895
Cryptosporidium parvum	species	5807
Cyclophyllidea	order	6201
Cystoisospora	genus	242060
Isospora belli	species	482538
Dibothriocephalus	genus	2267273
Dibothriocephalus latus	species	60516
Diphylobothrium latum	species	60516
Dicrocoeliidae	family	73421
Dicrocoelium	genus	57077
Dicrocoelium dendriticum	species	57078
Dientamoeba	genus	43351
Dientamoeba fragilis	species	43352
Dientamoebidae	family	740973
Digenea	subclass	6179
Digenea incertae sedis	noRank	757223
Diectophymatidae	family	513040
Diectophyme	genus	513042
Diectophyme renale	species	513045
Diphylobothriidae	family	28843
Diphylobothriidea	order	1224679
Diphylobothrium	genus	28844
Diplomonadida	order	5738
Dipylidiidae	family	109303
Dipylidium	genus	66786
Dipylidium caninum	species	66787
Dirofilaria	genus	6286
Discosea	noRank	555280
Dorylaimia	subclass	1457286
Dracunculidae	family	318477
Dracunculoidea	superfamily	55872
Dracunculus	genus	318478
Dracunculus medinensis	species	318479
Ecdysozoa	noRank	1206794
Echinococcus	genus	6209
Echinococcus granulosus	species	6210
Echinococcus granulosus granulosus	subspecies	1537665
Echinococcus granulosus group	species group	2212966

Echinostoma	genus	27847
Echinostoma hortense	species	48216
Echinostomata	suborder	27841
Echinostomatidae	family	99737
Echinostomatoidea	superfamily	404429
Eimeriorina	suborder	423054
Encephalitozoon	genus	6033
Encephalitozoon cuniculi	species	6035
Encephalitozoon hellem	species	27973
Encephalitozoon intestinalis	species	58839
Entamoeba nana	species	110788
Enoplea	class	119088
Entamoeba	genus	5758
Entamoeba coli	species	110766
Entamoeba dispar	species	46681
Entamoeba gingivitis	species	38877
Entamoeba hartmanni	species	110767
Entamoeba histolytica	species	5759
Entamoeba polecki	species	110769
Entamoebidae	family	33084
Enterobius	genus	51027
Enterobius vermicularis	species	51028
Enterocytozoon	genus	27971
Enterocytozoon bienewisi	species	31281
Enterocytozoonidae	family	27970
Enteromonadidae	family	306054
Enteromonas	genus	460188
Enteromonas hominis	species	460189
Eucestoda	subclass	6200
Eucoccidiorida	order	75739
Euglenozoa	noRank	33682
Eustrongylides	genus	513041
Fasciola	genus	6191
Fasciola hepatica	species	6192
Fasciolidae	family	27843
Fasciolopsis buski	species	27845
Filarioidea	superfamily	6295
Fornicata	noRank	207245
Fungi incertae sedis	noRank	112252
Gastrodiscoides	genus	408851
Gastrodiscoides hominis	species	408852
Giardia	genus	5740
Giardia ardeae	species	5743

<i>Giardia cati</i>	species	1631188
<i>Giardia duodenalis</i>	species	5741
<i>Giardia intestinalis</i>	species	5741
<i>Giardia lamblia</i>	species	5741
<i>Giardia microti</i>	species	61965
<i>Giardia muris</i>	species	5742
<i>Giardia peramelis</i>	species	1822422
<i>Giardia psittaci</i>	species	187323
Giardiinae	subfamily	68459
<i>Gongylonema</i>	genus	637852
Gorgoderoidea	superfamily	1776223
Haemosporida	order	5819
Heterolobosea	class	5752
Heterophyes	genus	104454
<i>Heterophyes heterophyes</i>	species	1849833
Heterophyidae	family	84526
Hexamitidae	family	5739
Hymenolepididae	family	6214
<i>Hymenolepis</i>	genus	6215
<i>Hymenolepis diminuta</i>	species	6216
<i>Hymenolepis nana</i>	species	102285
<i>Iodamoeba</i>	genus	1089792
<i>Iodamoeba</i> sp. RL1	species	1089794
<i>Iodamoeba</i> sp. RL2	species	1089793
Kinetoplastida	order	5653
<i>Leishmania</i>	subgenus	38568
<i>Leishmania</i>	genus	5658
<i>Leishmania aethiopica</i>	species	5667
<i>Leishmania aethiopica</i> species complex	species group	38570
<i>Leishmania amazonensis</i>	species	5659
<i>Leishmania braziliensis</i>	species	5660
<i>Leishmania chagasi</i>	species	44271
<i>Leishmania donovani</i>	species	5661
<i>Leishmania donovani</i> species complex	species group	38574
<i>Leishmania guyanensis</i>	species	5670
<i>Leishmania guyanensis</i> species complex	species group	38579
<i>Leishmania infantum</i>	species	5671
<i>Leishmania major</i>	species	5664
<i>Leishmania mexicana</i>	species	5665
<i>Leishmania mexicana</i> species complex	species group	38582
<i>Leishmania mexicana venezuelensis</i>	subspecies	155283
<i>Leishmania panamensis</i>	species	5679
<i>Leishmania tropica</i>	species	5666

Leishmania peruviana	species	5681
Leishmaniinae	subfamily	1286322
Loa	genus	7208
Loa loa	species	7209
Longamoebia	order	1485168
Macracanthorhynchus	genus	60533
Macracanthorhynchus hirudinaceus	species	1032456
Mammomonogamus	genus	1968513
Mammomonogamus laryngeus	species	2100468
Mansonella	genus	42230
Mansonella ozzardi	species	122354
Mansonella perstans	species	42231
Mansonella streptocerca	species	1761664
Metagonimus	genus	84527
Metagonimus yokogawai	species	84529
Microsporidia	phylum	6029
Moniliformida	order	10234
Moniliformidae	family	10235
Moniliformis	genus	10236
Moniliformis moniliformis	species	10237
Naegleria	genus	5761
Naegleria fowleri	species	5763
Nanophyetidae	family	138579
Nanophyetus	genus	197560
Nanophyetus salmincola	species	240278
Nematoda	phylum	6231
Neodiplostomidae	family	84100
Neodiplostomum	genus	84101
Neodiplostomum seoulense	species	84102
Oesophagostomum	genus	52564
Oligacanthorhynchida	order	60527
Oligacanthorhynchidae	family	60528
Onchocerca	genus	6281
Onchocerca volvulus	species	6282
Onchocercidae	family	6296
Opisthorchiata	suborder	6194
Opisthorchiida	order	6193
Opisthorchiidae	family	6196
Opisthorchis	genus	6197
Oxyuridae	family	51026
Oxyuridomorpha	infraorder	70426
Oxyuris	genus	132388
Oxyuris equi	species	132389

Oxyuroidea	superfamily	51025
Parabasalia	noRank	5719
Paracapillaria	genus	1457283
Capillaria philippinensis	species	1457282
Paragonimus	genus	34503
Paragonimus westermani	species	34504
Paramphistomidae	family	27852
Paramphistomoidea	superfamily	27850
Pentatrichomonas	genus	5727
Pentatrichomonas hominis	species	5728
Phaneropsolus	genus	1550885
Physaloptera	genus	75558
Physalopteridae	family	75557
Physalopteroidea	superfamily	331417
Piroplasmida	order	5863
Plagiorchiida	order	27871
Plasmodiidae	family	1639119
Plasmodium	genus	5820
Plasmodium falciparum	species	5833
Plasmodium malariae	species	5858
Plasmodium ovale	species	36330
Plasmodium vivax	species	5855
Platyhelminthes	phylum	6157
Pneumocystidaceae	family	44281
Pneumocystidales	order	37987
Pneumocystidomycetes	class	147553
Pneumocystis	genus	4753
Pneumocystis carinii	species	4754
Pronocephalata	suborder	1269772
Protostomia	noRank	33317
Pseudoterranova	genus	6270
Pseudoterranova decipiens	species	6271
Pygidiopsis	genus	104471
Pygidiopsis summa	species	104472
Retortamonadidae	family	193075
Retortamonas	genus	193076
Rhabditida	order	6236
Rhabditomorpha	infraorder	2301119
Sappinia	genus	343528
Sappinia diploidea	species	343529
Sarcocystidae	family	5809
Sarcocystis	genus	5812
Sarcocystis cruzi	species	5817

Schistosoma	genus	6181
Schistosoma bovis	species	6184
Schistosoma haematobium	species	6185
Schistosoma intercalatum	species	6187
Schistosoma japonicum	species	6182
Schistosoma mansoni	species	6183
Schistosoma mekongi	species	38744
Schistosomatidae	family	31245
Schistosomatoidea	superfamily	31244
Schizopyrenida	order	41165
Spirometra	genus	46580
Spirometra mansonioides	species	46899
Spirurina	suborder	6274
Spiruromorpha	infraorder	2072716
Strongyloidea	superfamily	27829
Strongyloides	genus	6247
Strongyloides fuelleborni	species	44441
Strongyloides stercoralis	species	6248
Strongyloididae	family	6246
Syngamidae	family	120855
Taenia	genus	6202
Taenia asiatica	species	60517
Taenia saginata	species	6206
Taenia solium	species	6204
Taeniidae	family	6208
Taphrinomycotina	sybphylum	451866
Ternidens	genus	317881
Ternidens deminutus	species	317882
Thecamoebida	family	242543
Theileria	genus	5873
Theileria equi	species	5872
Theileriidae	family	27994
Toxascaris	genus	59263
Toxascaris leonina	species	59264
Toxocara	genus	6264
Toxocara canis	species	6265
Toxocara cati	species	6266
Toxocaridae	family	33259
Toxoplasma	genus	5810
Toxoplasma gondii	species	5811
Trematoda	class	6178
Trichinella	genus	6333
Trichinella pseudospiralis	species	6337

Trichinella spiralis	species	6334
Trichinellida	order	6329
Trichinellidae	family	6332
Trichomonadida	order	37104
Trichomonadidae	family	181550
Trichomonas	genus	5721
Trichomonas vaginalis	species	5722
Trichostrongylus	genus	6318
Trichuridae	family	119093
Trichuris	genus	36086
Trichuris muris	species	70415
Trichuris ovis	species	93034
Trichuris suis	species	68888
Trichuris trichiura	species	36087
Trichuris vulpis	species	219738
Tritrichomonadida	order	740972
Troglotremata	suborder	116925
Troglotrematidae	family	34502
Trypanosoma	genus	5690
Trypanosoma brucei gambiense	subspecies	31285
Trypanosoma brucei rhodesiense	subspecies	31286
Trypanosoma cruzi	species	5693
Trypanosoma rangeli	species	5698
Trypanosomatidae	family	5654
Unikaryonidae	family	36734
Vahlkampfiidae	family	5765
Viannia	subgenus	37616
Wuchereria	genus	6292
Wuchereria bancrofti	species	6293
Xiphidiata	suborder	27872

7. Discussion

NGS technologies have provided unprecedented access and flexibility to archaeogeneticists, and this is exemplified by the projects included in this dissertation. A ubiquitous problem when such wealth of data is made available is determining how best to use it. As archaeogenetics develops as a discipline, and the pace of “first discoveries” slows, deeper explorations of existing datasets will be a rich source of information that can be used to, for example, develop the evolutionary stories of single species or complexes of bacteria, or refine our understanding of how microbiota preserve. Below, the ways in which the projects presented in this dissertation advance disciplinary understanding on multiple levels – both topical and methodological – are discussed. The existing limitations to further development and knowledge growth in microbial archaeogenetics are also discussed.

7.1 Challenges in uncovering the origins of tuberculosis

In Manuscript A, we report the successful recovery of MTBC DNA from a calcified lung nodule from the mummified remains of Bishop Peder Winstrup of Lund, the efficient enrichment of that DNA, and the utilization of the resulting genome as a calibration point with which to date the MTBC. We confirmed the previous findings of Bos and colleagues (Bos et al., 2014) that the extant MTBC is likely less than 6000 years old using a variety of models, including the birth-death skyline multi-rho tree prior (Stadler et al., 2013), which had the most favorable likelihood when compared with all other models we tried. Our results are fundamentally due to the excellent preservation of DNA in Bishop Winstrup’s calcified lung nodule as a preservation environment for DNA, which allowed us to bypass many of the complications inherent in studying tuberculosis from archaeological contexts. To take the next steps in uncovering the origins of tuberculosis and reconciling the disparities between different lines of evidence for the disease’s antiquity (or lack thereof) there are many challenges to consider and prepare for.

7.1.1 Archaeological sources of MTBC DNA

Dental pulp chambers are the most systematically targeted sampling areas in ancient pathogen research due to their high vascularization. The chamber is protected from excess degradation by dental enamel, and would have likely been exposed to any pathogen present in high volume in the blood stream at the individual’s time of death. For example, pulp chambers have been shown to be excellent harbors of *Yersinia pestis* (Bos et al., 2011; Drancourt et al., 1998; Feldman et al., 2016; Spyrou et al., 2018b). However, no publications have reported ancient and verifiable MTBC DNA from the pulp chamber. This is not surprising, as tuberculosis rarely manifests as a septic condition (Landouzy septicemia), and

usually does so among immunocompromised patients or infants with congenital tuberculosis (Angoulvant et al., 1999; Geiss et al., 2005; Mazade et al., 2001; Michel et al., 2002; Pène et al., 2001). The most traditional and infectious manifestation of tuberculosis is pulmonary, and DNA from MTBC organisms has been detected by shotgun sequencing in modern sputum samples (Doughty et al., 2014). Intuitively, we would then turn to dental calculus as a promising preservation environment for tuberculosis due to its excellence as a proxy for saliva-borne bacteria and the human oral microbiome (Warinner et al., 2014, 2014). However, no study has presented convincing evidence of DNA from the *Mycobacterium tuberculosis* complex (MTBC) trapped in ancient dental calculus.

Ancient MTBC DNA has been identified with next generation sequencing (NGS) techniques from samples from vertebral bodies, ribs, mummified lung tissue and a calcified lung nodule (Bos et al., 2014; Chan et al., 2013; Kay et al., 2015; Warinner et al., 2014). Calcified nodules from human remains, frequently indicative of an infective process (Brown et al., 1994), based on the results presented in Manuscript A and two other studies appear to be excellent sources of DNA for the infecting pathogen and the host (Devault et al., 2017; Kay et al., 2014). In the case of tuberculosis, the exceptionally high proportion of host DNA could be in part due to the immune reaction that contributes to granuloma formation, in which non-infected macrophages and lymphocytes surround macrophages that have been invaded by bacilli (Brown et al., 1994; Russell, 2007). The size of calcified nodules implicated in tuberculosis infections complicate their recovery from archaeological contexts. In their *Brucella melitensis* study, Kay and colleagues extracted DNA from one of thirty-two calcified nodules ranging from ~0.6 to 2.2 cm in diameter collected from the pelvic girdle of skeletonized remains (2014). Devault and colleagues extracted DNA from *Staphylococcus saprophyticus* and *Gardnerella vaginalis* from two calcified nodules between 2 and 3 cm in diameter recovered from a single, stone-lined grave (2017). To my knowledge, nodules diagnosed as resulting from tuberculosis in archaeological remains have only come from mummified remains (Kim et al., 2016; Lunardini et al., 2012; Salo et al., 1994). Calcifications comparable in size to the one presented in Manuscript A (~5 mm) may easily be missed during excavations of skeletonized remains depending on the depositional context. Vertebral bodies and ribs, which often bear skeletal changes indicative of tuberculosis and are popular sites for lesion development due to their high oxygenation (Roberts and Buikstra, 2003), do not always lead to successful identification of MTBC DNA even when signs are visible. However, they are far more accessible and likely to preserve in and be recovered from common burial contexts. The literature thus far does not suggest a universally reliable source of ancient tuberculosis DNA.

7.1.2 *Authentication and environmental mycobacteria*

A relatively unexplored complication of any study of ancient MTBC biomolecules is the seeming omnipresence of environmental mycobacteria, which share close phylogenetic relationships with MTBC species. Environmental mycobacteria have previously been discussed as a complication of genetic studies of the MTBC (Warinner et al., 2017; Wilbur et al., 2009). The commonly used genetic markers for PCR amplification of tuberculous mycobacteria, IS6110 and IS1081, were tested for modern clinical and epidemiological contexts (Behr and Small, 1997; Mathuria et al., 2008). To my knowledge, no systematic testing has been performed evaluating the specificity of IS6110 and IS1081 amplifications in archaeological contexts or non-archaeological soils in which non-tuberculous mycobacteria are abundant (Falkinham, 2015; Hruska and Kaevska, 2013), though one study reported the amplification of sequences that were highly similar, but not identical, to IS6110 from archaeological bone (Müller et al., 2015). Thus, identification of MTBC DNA at the genomic level is preferable not only for downstream utilization of the data to answer questions about MTBC evolution, but also for authenticating the DNA as being ancient and from a tuberculosis-causing organism. Preliminary identification and authentication using a lowest common ancestor (LCA) taxonomic binning algorithm as presented in Manuscript A, or a competitive mapping approach (e.g. Andrades Valtueña et al., 2017), in which environmental mycobacteria and the presence, absence, and degree of damage can be considered through nucleotide misincorporation plots, provides greater confidence and decreases the likelihood of making a false or unverifiable claim.

7.1.3 *Challenges of demographic inference and molecular dating using ancient DNA*

The impact of the tree prior on posterior estimates in Bayesian phylogenies has been demonstrated with empirical and *in silico* data (Boskova et al., 2018; Möller et al., 2018). Past estimates of the time to the MRCA and population dynamics of the MTBC have depended on models built on the Kingman coalescent (Bos et al., 2014; Brynildsrud et al., 2018; Kay et al., 2015; O'Neill et al., 2019), which describes the relationship of lineages moving backwards through time. However, the appropriateness of the Kingman coalescent for inferring the population dynamics of tuberculosis from time-series data is questionable. In Manuscript A, multiple models are used for Bayesian estimation of the time to the MRCA including those based on the coalescent and an alternative model base: the forward-in-time birth-death process.

Where phylodynamics are concerned, the coalescent-based Bayesian skyline does not take into account epidemiological parameters such as incidence and prevalence, and the coalescent approaches currently available in BEAST2 assume a random sample of uniform proportion from a much larger population (Stadler et al., 2013). Though the birth-death skyline model was developed with pathogen outbreak circumstances in mind, in which

researchers could confidently estimate the proportion of cases represented in the analysis due to near-complete sampling in emergent situations, it allows for the parameterization of different sampling proportions or *rho* priors (sampling events with a given proportion that occur at distinct time points) over time. In theory, this makes birth-death skyline models ideal for MRCA and phylodynamic estimates using ancient pathogen DNA, as the sampling rate of genomes is uneven over time. However, the flexibility of birth-death skyline models also makes them more complicated to parameterize correctly. In Manuscript A, we select relatively arbitrary *rho* priors, as we could not at the time of the study develop a methodology to estimate a reasonable *rho* prior for our ancient samples. This would require estimation of the total MTBC population (or lineage diversity) at the sampling time of each ancient genome utilized. The development of a model or methodology for approximating total diversity for ancient samples (not only for tuberculosis, but also for other pathogens), that preferably is not circular and does not depend solely on previous skyline estimates, would be useful for increasing the accuracy of phylodynamic estimates utilizing ancient DNA in the future. This issue will increase in relevance as more ancient pathogen genomes are discovered and made publicly available.

Another limitation of existing phylodynamic methodologies is the assumption of bifurcating trees. For instance, the Kingman coalescent assumes only two lineages can coalesce at a time (Irwin et al., 2016), and many types of organism violate this assumption (Tellier and Lemaire, 2014). The MTBC and Lineage 4 phylogenies appear to present highly clonal, “starburst” topologies that also violate the pairwise coalescence assumption. Development and testing of a package for BEAST2 that explicitly accommodates phylogenies with polytomies warrants further investigation.

In Manuscript A, we used multiple tree priors to estimate the most recent common ancestor of the MTBC and Lineage 4 of the MTBC using all available ancient complex genomes of sufficient quality for calibration of tips. Our experimentation with multiple models and our model selection by stepping-stone sampling sought to provide the best estimates possible given our data and the packages and models currently available in BEAST2 (Bouckaert et al., 2014; Xie et al., 2011). Of the models in our study, the birth death skyline model performs best for both datasets (Stadler et al., 2013).

The temporal structure in our datasets as calculated with TempEst (Rambaut et al., 2016) is present but not particularly strong, especially for the L4 dataset. The root-to-tip distance of each genome is largely determined by the lineage or sublineage to which the isolate belongs. The strength of our MRCA estimates is bolstered by the relative agreement between all tested models. However, our analysis would be further strengthened by a date randomization test, in which the calibrated tip dates for ancient genomes are randomly mixed among the modern genomes (Ramsden et al., 2009). This aids in supporting or rebuking the

clock-like behavior of the data within the model, and has been used with ancient calibration points in prior publications (Ross et al., 2018; Wagner et al., 2014). Meaningfulness of the data is evaluated by whether or not the mean substitution rate estimates of the randomizations lie within the 95% highest posterior density (HPD) interval of a true model or whether the 95% HPD intervals of the randomizations overlap at all with that of the true model, depending on the desired level of stringency (Duchêne et al., 2015).

7.1.4 Where to aim for future studies

Though almost any ancient MTBC genome non-identical to those already sequenced would add to our body of knowledge about the history of tuberculosis, the most compelling and informative genomes would be those that are basal to one or more lineages of the MTBC. All published ancient genomes sit firmly within known diversity of extant strains, and though they offer temporal calibration, their phylogenetic positions do not offer revelations regarding the movement or evolution of tuberculosis over space and time. Where and for what types of samples could one look to uncover more basal strains of the MTBC?

As discussed in Manuscript A, ancient individuals with signs of tuberculosis from sub-Saharan Africa pre-dating 1000 years BP would offer the best chance of uncovering a basal genome. Until recently, sampling human remains for preserved pathogens would have seemed like a fool's errand given the unfavorable climates for DNA preservation over most of the continent. However, recent studies have proven that sub-Saharan African (Llorente et al., 2015), extremely arid environments (Loosdrecht et al., 2018), and tropical environments (Posth et al., 2018) in general do not preclude the recovery of ancient DNA. Thus far, mummified or partially mummified remains have seemed to be the best preservation environments for MTBC DNA. As demonstrated by Manuscript A, the best burial context for the preservation of ancient MTBC DNA is a crypt interment, or one otherwise removed from direct contact with soil. Among analyses using NGS techniques to identify MTBC DNA, all samples came from individuals with indications of tuberculosis infection present in the remains. In the case of the Hungarian mummies analyzed by Chan and colleagues (Chan et al., 2013) and Kay and colleagues (Kay et al., 2015), and in the case of Manuscript A, there were also historical records supporting a diagnosis of tuberculosis. Similar historical situations would be the most promising contexts for investigating ancient tuberculosis, but this would likely not yield the basal sequences that would be most informative for the history of the disease.

Though increasing the risk of great expenditure for negative results, continued shotgun screening of remains with skeletal indicators of tuberculosis is likely the best avenue forward for uncovering more basal ancient strains. Archaeogeneticists may form alternative hypotheses pertaining to the human DNA or other microbial DNA from the remains they choose to test to mitigate the financial and professional burden of negative results. As shown

by the projects in this dissertation, NGS data can be used in many valuable ways. Also, a tool such as HOPS (Huebler et al., 2019) makes it easier to screen shotgun sequencing data for a plethora of pathogens, and if utilized regularly by researchers producing such data, will reduce the likelihood that an ancient genetic MTBC signal will be missed when the DNA of the host or a different microbe is the target. For this reason, it would be beneficial to the field of archaeogenetics as a whole if shotgun sequencing were the standard course of action, as opposed to targeted amplification or enrichment prior to generalized screening. The next paradigm-shifting discovery pertaining to the evolutionary history of tuberculosis may be opportunistic, and stem from a completely unexpected context.

7.2 Characterizing ancient human microbiota

In Manuscripts B and C, we explore genetic preservation potential in three types of archaeological material – dentin, calculus, and latrine sediment – using shotgun-sequenced data. The calculus and latrine sediments analyzed in this thesis consistently preserve traces of the individual oral and collective gut microbiota, respectively, and we found a subset of our dentin samples preserved an oral microbiome signal as well.

7.2.1 *Facilitating qualitative interpretability of large metagenomic datasets*

A forefront challenge in characterizing ancient human microbiota (or any ancient microbial community surpassing a dozen or so taxa) is efficiently and simultaneously authenticating and sorting the genetic information from the noisy, taxonomically diverse background of most archaeological specimens. Manuscript B made great strides in testing the assumptions regarding and limitations on DNA preservation in dental calculus, the most dependable known proxy for any ancient endogenous human microbiome. However, our two methods of summarizing the oral microbiome content of the dentin and calculus samples included in our study did not offer a means of ancient DNA authentication by assessing damage or to identify oral microbiome content on a species-by-species basis in an *efficient* manner. The bottom-up approach we employed (as opposed to the top-down Bayesian SourceTracker2 approach) consisted of grouping identified species in each sample by their most common environment and manifestation based on literature surveys. It is not scalable to larger datasets, barring the development and curation of a meta-database specifically designed for this purpose, and would still require additional manual assessment of DNA damage for alignments to each species. HOPS, which was still under development at the time Manuscript B was published, offers a partial solution to the scalability problem in its automated detection of damaged reads (Huebler et al., 2019). However, HOPS was designed to help automate the detection and authentication of individual pathogen species, and tested with a specific set of bacterial pathogens. The approach employed in Manuscript B, in which individual species are sorted into layered metadata categories is currently, to my

knowledge, not possible to replicate in an automated fashion with an existing tool or database.

7.2.2 *Limitations of modern databases and taxonomic classifiers*

There are additional problems that affect both modern and ancient microbiome characterization. A central problem is the scope of modern databases. Characterization of modern human microbiota, especially prior to the implementation of NGS techniques, has been hampered by difficulties in identifying unculturable microbes (Verma et al., 2018). The expanded Human Oral Microbiome Database currently contains 770 species, 57% of which are named, 13% of which are unnamed and cultivated, and 30% of which are uncultivated, unnamed phylotypes (Escapa et al., 2018). Similarly, the database we used to summarize the gut microbiome signal from the latrine sediments presented in Manuscript C is limited to cultured gut bacteria (Forster et al., 2019). Recent metagenomics efforts have dramatically increased representation of uncultured bacteria from human microbiota (Almeida et al., 2019; Pasolli et al., 2019), but the curation and classification pipeline for the new metagenomically assembled genomes may slow our ability to use them.

The study of ancient microbiomes is also hindered by the limitations of available taxonomic classifiers. Velsko and colleagues explored this issue specifically for ancient data by simulating an ancient metagenomic dataset with known composition and testing the ability of multiple taxonomic classifiers to identify all taxa correctly (Velsko et al., 2018). There was no obvious choice, and the authors concluded that the best taxonomic classifier for any project depended on the specific questions being asked.

All manuscripts included in this thesis use MALT (Vågene et al., 2018). MALT was used as the taxonomic classifier for the manuscripts for multiple reasons. MALT, when used in conjunction with the full nucleotide database from NCBI, is as close to a truly metagenomic tool as one can get, in the sense that it aligns the whole genetic content of a file with all sequences and regions available in the database instead of depending on marker genes for classification. In addition, it can resolve reads to the species or subspecies level if the reads are sufficiently specific. There is also a level of convenience to the use of MALT, as it is now integrated with HOPS, which automates the identification of organisms of interest and can extract important linked information, such as the fragment lengths of reads aligning to a given taxa, or signs of DNA damage in the form of nucleotide misincorporation (Huebler et al., 2019). A prototypical part of what has become HOPS was used for exactly that purpose in Manuscript B, and the HOPS pipeline was used in its full, current form for Manuscript C.

7.2.3 *Inferring microbiome dynamics and “normal” microbiomes*

After overcoming the hurdle of accurately identifying microbes known to be part of human microbiota, there comes the next level of interpreting microbial community dynamics from what species are present and absent and perhaps, if datasets can be properly

normalized, the abundances of different species. Making accurate associations between health states and the presence or absence of particular taxa is challenging and misleading in modern populations (Menon et al., 2018). Efforts to characterize “normal,” healthy human microbiomes like the Human Microbiome Project (Consortium, 2012; Group et al., 2009) are limited by geographical and temporal reach. Ancient microbiome data can provide nuance to what is considered healthy and not. For example, the “Red Complex” of oral bacteria was named such for its association with severe periodontal disease (Socransky et al., 1998). However, all three members of the “Red Complex” (*Tannerella forsythia*, *Treponema denticola*, and *Porphyromonas gingivalis*) were found in all individuals in the regional dataset presented in Manuscript B. Severe periodontitis, which can be identified in skeletonized human remains, was not identified in all individuals in the dataset. All three were also found in the study of medieval calculus from Germany by Warinner and colleagues (2014). This leads us to rethink what we mean when we call certain microbes “pathogens.” Perhaps it is not the presence or absence of the “Red Complex” that determines oral health, but the dynamic relationship of those organisms to other microbes in the oral cavity.

Microbial archaeogenetics can also provide temporal context to the human gut microbiome. Studies in modern individuals have scratched the surface of the effects of industrialized life on the microbial composition of the gut, discovering for instance, that the gut microbiota of people living more traditional or non-industrialized lifestyles (e.g. not using antibiotics or consuming processed foods) contain *Treponema* diversity, while those of more industrialized people do not (Angelakis et al., 2019; Obregon-Tito et al., 2015). As further similar studies are released, making use of metagenomics to detect unculturable microbes (Almeida et al., 2019; Pasolli et al., 2019), the resulting information can be used to interpret ancient gut microbiota and recursively reinterpret modern gut microbiota. It is important to remember that modern non-industrialized populations are not fossils. Though it would be extremely difficult to detect and authenticate any microbes unique to ancient gut microbiota, we would be able to detect shifts in community composition among known gut-associated microbes, and perform in-depth analyses of gut-associated microbes we find in the past to trace organism evolution. As we learn more about modern microbes, we will learn more about ancient microbes, which will then allow us to reframe what we consider to be “normal” microbiota.

7.2.4 *Dentin as a potential source for oral microbes*

One of the unexpected findings presented in Manuscript B was the presence of an oral microbiome signal in DNA from pulp chamber dentin. Though the majority of microbial DNA was environmental, a portion in some dentin samples was from oral microbes (up to 20%). As discussed in the manuscript, this presents an opportunity to investigate individual oral microbes in cases when dental calculus may be scarce or unavailable, such as with early or

archaic humans (Warinner et al., 2014). The stochasticity of oral microbe preservation in dentin could be due to individual circumstances of decomposition. To my knowledge, there has been no metagenomic exploration of post-mortem microbial communities inside teeth, though there has been at least one project investigating the post-mortem metagenomics of the human oral cavity (Hyde et al., 2013).

7.3 Challenges in studying ancient eukaryotic parasites with ancient DNA

Human-infecting eukaryotic parasites offer a proxy for studying human behavioral and biological evolution (Mitchell, 2013; Perry, 2014). As stated in the introduction and Manuscript C, such ancient pathogens have been mostly studied using microscopy to identify parasite eggs. In the two studies on the bulk latrine sediments from Riga and Jerusalem, microscopy and immunoassays were used in tandem (Yeh et al., 2015, 2014). Those two methodologies are sufficient for identifying the presence of parasites at the genus or species level, but ideally, genetics could offer more.

In Manuscript C, DNA was extracted and analyzed differently than in prior publications presenting genetic evidence of parasites from latrine sediments or paleofeces. Instead of extracting DNA from sieved, concentrated parasite eggs, the DNA was extracted from relatively unadulterated portions of the bulk sediment from each site, without any knowledge of whether or not eggs were included directly in the extracted aliquots. Though further studies are required to test this, it is possible archaeogenetics can provide evidence of eukaryotic parasites abundant in a deposit without directly extracting an egg or the organism itself in the case of protozoa. As the egg decomposes, DNA may be released into the surrounding sediments and bind to them (Slon et al., 2017; Vandeventer et al., 2012). If this can be shown to be the case, genetics would provide a possibly more representative view than microscopy of what parasites were present in a latrine pit if sufficiently-spaced samples were taken. In theory, genetics would also offer information that would allow researchers to identify parasites to the species, subspecies, or lineage level, and perhaps when enough ancient data is available, perform population genetics analyses on the parasites in question, such as when Søe and colleagues analyzed mitochondrial haplotypes of *Ascaris lumbricoides* and *Trichuris trichiura* (2018). However, barriers to making full use of our access to genetic sequences from ancient parasites remain.

Genome sizes among eukaryotic parasites vary considerably. Though eukaryotic parasites fall under the same thematic category, they are a taxonomically diverse group, including macroscopic worms and single-celled protozoans. For example, nematode and platyhelminth parasites range in genome size from 42 to 1,295 Mb (International Helminth Genomes Consortium, 2019). These large sizes (as compared to bacterial and archaeal

genomes) contribute to difficulties in the reconstruction of complete genomes from ancient data. Thus far, researchers have only been able to reassemble mitochondrial genomes of three ancient parasites, and not always completely (Marciniak et al., 2016; S e et al., 2018). To effectively move past the presence-absence stage of analysis, it is likely a capture assay will be necessary to extract meaningful data out of any ancient dataset. However, to properly design such an assay and evaluate its results, high-quality reference sequences must be available.

As discussed in Manuscript C, the paucity of modern reference data for many important species (e.g. *Ascaris lumbricoides*) currently limits the ability of researchers to study ancient genomes from these organisms. A recent publication from the International Helminth Consortium has greatly expanded the number of publicly available genome-level assemblies, specifically for helminths. One such assembly acted as the reference genome for the *Ascaris lumbricoides* sequences presented in Manuscript C, but the lack of genome-wide sequence data for organisms such as *Oxyuris equi* (the equid pinworm, previously identified via microscopy in the Riga latrine) and *Diphyllobothrium dendriticum* (a close relative of *Dibothriocephalus latus*, and one which may be confused for *D. latus* in microscopy investigations) limited the questions Manuscript C was able to address.

Moving forward, it would be valuable to researchers studying ancient eukaryotic parasites for there to be more abundant reference data. In addition, however, it would be valuable to test how the genomic sequences that are currently available behave as simulated ancient DNA. An analysis of what taxonomic categorization tools work best with simulated ancient eukaryotic parasite data, such as that presented by Velsko and colleagues for bacterial studies (2018), or a replication of the experiment investigating simulated ancient DNA read specificity on a taxon-by-taxon basis among bacterial pathogens presented in the HOPS manuscript (Huebler et al., 2019), would set helpful standards for the future. ///

7.4 Conclusion

The projects presented in this thesis are focused on data extracted from diverse archaeological materials that pose unique challenges and opportunities. Additionally, each project advances the field of ancient microbial genomics in different ways. In Manuscript A, an unprecedentedly clear signal for MTBC DNA was extracted from a calcified lung nodule. Not only did this provide genetic confirmation that the individual in question had been infected with tuberculosis, but it provided an example of the utility of calcified lung nodules in studying ancient tuberculosis (and host genetics), a high quality genome to add to the growing dataset of ancient MTBC sequences that aid us in describing past pathogen diversity, a new calibration point with which to date the emergence of the MTBC, and an opportunity to test new models for describing the phylogenetic history of the MTBC.

Manuscript B broadens this dissertation in scope from a single pathogen to the whole microbial communities found in dentin and dental calculus. In doing so it provides the first systematic, large volume study evaluating the preservation potential of dental calculus in comparison to dentin, the more traditional ancient DNA source. With the publication of Manuscript B, it was established that dental calculus consistently provides more raw DNA than paired dentin, and based on metagenomic analysis the microbial portion of DNA mostly stems from the oral microbiome. Ecological measurements show this signal dominating any environmental microbial signal. Based on Manuscript B, there now exists a baseline reference dataset defining standard expectations for DNA preservation in dental calculus. Manuscript C also takes a broad metagenomic approach to an archaeological proxy for human microbiota and pathogens, as the first ancient DNA analysis of bulk latrine sediments assessing not only the presence of eukaryotic parasites, but preservation of a community gut microbiome signal. This is a first foray into metagenomic investigations of a highly abundant archaeological material.

Microbial archaeogenetics as a field brings attention to neglected and re-emerging diseases with long and potentially co-evolutionary histories with *Homo sapiens*, and contextualizes human microbiota through the lens of time. Applying modern methodologies to understand past health, including the epidemiology and genetic makeup of ancient pathogens as well as commensal microbes, could offer valuable insights into modern health. Research into ancient pathogens and microbiota is still young and developing, with new tools streamlining the screening of ancient metagenomic data and opening doors to unexpected findings.

8. References

- Achtman, M., 2016. How old are bacterial pathogens? *Proc. R. Soc. B* 283, 20160990. <https://doi.org/10.1098/rspb.2016.0990>
- Adler, C.J., Dobney, K., Weyrich, L.S., Kaidonis, J., Walker, A.W., Haak, W., Bradshaw, C.J.A., Townsend, G., Sołtysiak, A., Alt, K.W., Parkhill, J., Cooper, A., 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics* 45, 450–455. <https://doi.org/10.1038/ng.2536>
- Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., Finn, R.D., 2019. A new genomic blueprint of the human gut microbiota. *Nature* 568, 499. <https://doi.org/10.1038/s41586-019-0965-1>
- Anderson, T.J.C., 2001. The dangers of using single locus markers in parasite epidemiology: *Ascaris* as a case study. *Trends in Parasitology* 17, 183–188. [https://doi.org/10.1016/S1471-4922\(00\)01944-9](https://doi.org/10.1016/S1471-4922(00)01944-9)
- Andrades Valtueña, A., Mitnik, A., Key, F.M., Haak, W., Allmäe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., Massy, K., Novak, M., Pfrengle, S., Reinhold, S., Šlaus, M., Spyrou, M.A., Szécsényi-Nagy, A., Törv, M., Hansen, S., Bos, K.I., Stockhammer, P.W., Herbig, A., Krause, J., 2017. The Stone Age Plague and Its Persistence in Eurasia. *Current Biology*. <https://doi.org/10.1016/j.cub.2017.10.025>
- Angelakis, E., Bachar, D., Yasir, M., Musso, D., Djossou, F., Gaborit, B., Brah, S., Diallo, A., Ndombe, G.M., Mediannikov, O., Robert, C., Azhar, E.I., Bibi, F., Nsana, N.S., Parra, H.-J., Akiana, J., Sokhna, C., Davoust, B., Dutour, A., Raoult, D., 2019. *Treponema* species enrich the gut microbiota of traditional rural populations but are absent from urban individuals. *New Microbes and New Infections* 27, 14–21. <https://doi.org/10.1016/j.nmni.2018.10.009>
- Angoulvant, D., Mohammadi, I., Duperret, S., Bouletreau, P., 1999. Septic shock caused by *Mycobacterium tuberculosis* in a non-HIV patient. *Intensive Care Med* 25, 238.
- Araújo, A., Ferreira, L.F., 2000. Paleoparasitology and the antiquity of human host-parasite relationships. *Mem. Inst. Oswaldo Cruz* 95, 89–93. <https://doi.org/10.1590/S0074-02762000000700016>
- Araújo, A., Reinhard, K., Ferreira, L.F., Pucu, E., Chieffi, P.P., 2013. Paleoparasitology: the origin of human parasites. *Arq. Neuro-Psiquiatr.* 71, 722–726. <https://doi.org/10.1590/0004-282X20130159>
- Araújo, A., Reinhard, K.J., Ferreira, L.F., Gardner, S.L., 2008. Parasites as probes for prehistoric human migrations? *Trends in Parasitology* 24, 112–115. <https://doi.org/10.1016/j.pt.2007.11.007>
- Armelagos, G.J., Brown, P.J., Turner, B., 2005. Evolutionary, historical and political economic perspectives on health and disease. *Social Science & Medicine, The Social Production of Health: Critical Contributions from Evolutionary, Biological and Cultural Anthropology: Papers in Memory of Arthur J. Rubel* The Social Production of Health: Critical Contributions from Evolutionary, Biological and Cultural Anthropology: Papers in Memory of Arthur J. Rubel 61, 755–765. <https://doi.org/10.1016/j.socscimed.2004.08.066>
- Arriaza, B.T., Salo, W., Aufderheide, A.C., Holcomb, T.A., 1995. Pre-Columbian tuberculosis in northern Chile: molecular and skeletal evidence. *Am. J. Phys. Anthropol.* 98, 37–45. <https://doi.org/10.1002/ajpa.1330980104>
- Arrizabalaga, J., 2002. Problematizing Retrospective Diagnosis in the History of Disease. *Asclepio* 54, 51–70.
- Ascaris lumbricoides* - WormBase ParaSite [WWW Document], n.d. URL https://parasite.wormbase.org/Ascaris_lumbricoides_prjeb4950/Info/Index (accessed 5.20.19).
- Ash, L.R., Orihel, T.C., 2015. *Ash & Orihel's Atlas of Human Parasitology*, 5th ed. American Society for Clinical Pathology Press.

- Aurrecoechea, C., Barreto, A., Brestelli, J., Brunk, B.P., Cade, S., Doherty, R., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Hu, S., Iodice, J., Kissinger, J.C., Kraemer, E.T., Li, W., Pinney, D.F., Pitts, B., Roos, D.S., Srinivasamoorthy, G., Stoeckert, C.J., Wang, H., Warrenfeltz, S., 2013. EuPathDB: The Eukaryotic Pathogen database. *Nucl. Acids Res.* 41, D684–D691. <https://doi.org/10.1093/nar/gks1113>
- Auwerda, G.A.V. der, Carneiro, M.O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* 43, 11.10.1-11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Bahr, N.C., Antinori, S., Wheat, L.J., Sarosi, G.A., 2015. Histoplasmosis infections worldwide: thinking outside of the Ohio River valley. *Curr Trop Med Rep* 2, 70–80. <https://doi.org/10.1007/s40475-015-0044-0>
- Baker, O., Lee, O.Y.-C., Wu, H.H.T., Besra, G.S., Minnikin, D.E., Llewellyn, G., Williams, C.M., Maixner, F., O’Sullivan, N., Zink, A., Chamel, B., Khawam, R., Coqueugniot, E., Helmer, D., Le Mort, F., Perrin, P., Gourichon, L., Dutailly, B., Pálfi, G., Coqueugniot, H., Dutour, O., 2015. Human tuberculosis predates domestication in ancient Syria. *Tuberculosis, Supplement issue: Tuberculosis in Evolution* 95, S4–S12. <https://doi.org/10.1016/j.tube.2015.02.001>
- Beck, J.D., Offenbacher, S., 2005. Systemic Effects of Periodontitis: Epidemiology of Periodontal Disease and Cardiovascular Disease. *Journal of Periodontology* 76, 2089–2100. <https://doi.org/10.1902/jop.2005.76.11-S.2089>
- Behr, M.A., Small, P.M., 1997. Molecular Fingerprinting of *Mycobacterium tuberculosis*: How Can It Help the Clinician? *CLIN INFECT DIS* 25, 806–810. <https://doi.org/10.1086/515550>
- Blakely, R.L., Detweiler-Blakely, B., 1989. The Impact of European Diseases in the Sixteenth-Century Southeast: A Case Study. *Midcontinental Journal of Archaeology* 14, 62–89.
- Boritsch, E.C., Khanna, V., Pawlik, A., Honoré, N., Navas, V.H., Ma, L., Bouchier, C., Seemann, T., Supply, P., Stinear, T.P., Brosch, R., 2016. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. *PNAS* 201604921. <https://doi.org/10.1073/pnas.1604921113>
- Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514, 494–497. <https://doi.org/10.1038/nature13591>
- Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M., 2016. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* 5, e12994.
- Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., Wood, J., Earn, D.J.D., Herring, D.A., Bauer, P., Poinar, H.N., Krause, J., 2011. A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506–510. <https://doi.org/10.1038/nature10549>
- Boskova, V., Stadler, T., Magnus, C., 2018. The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. *Virus Evol* 4. <https://doi.org/10.1093/ve/vex044>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* 10, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>

- Braun, M., Collins Cook, D., Pfeiffer, S., 1998. DNA from Mycobacterium tuberculosis Complex Identified in North American, Pre-Columbian Human Skeletal Remains. *Journal of Archaeological Science* 25, 271–277. <https://doi.org/10.1006/jasc.1997.0240>
- Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *PNAS* 104, 14616–14621. <https://doi.org/10.1073/pnas.0704665104>
- Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., Pääbo, S., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38, e87–e87. <https://doi.org/10.1093/nar/gkp1163>
- Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., Soolingen, D. van, Cole, S.T., 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *PNAS* 99, 3684–3689. <https://doi.org/10.1073/pnas.052548299>
- Brown, K., Mund, D.F., Aberle, D.R., Batra, P., Young, D.A., 1994. Intrathoracic calcifications: radiographic features and differential diagnoses. *RadioGraphics* 14, 1247–1261. <https://doi.org/10.1148/radiographics.14.6.7855339>
- Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J., Alfsnes, K., Pettersson, J.O.-H., Kirkeleite, I., Fandinho, F., Silva, M.A. da, Perdigao, J., Portugal, I., Viveiros, M., Clark, T., Caws, M., Dunstan, S., Thai, P.V.K., Lopez, B., Ritacco, V., Kitchen, A., Brown, T.S., Soolingen, D. van, O'Neill, M.B., Holt, K.E., Feil, E.J., Mathema, B., Balloux, F., Eldholm, V., 2018. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. *Science Advances* 4, eaat5869. <https://doi.org/10.1126/sciadv.aat5869>
- Burbano, H.A., Hodges, E., Green, R.E., Briggs, A.W., Krause, J., Meyer, M., Good, J.M., Maricic, T., Johnson, P.L.F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F.W., Rasilla, M. de la, Fortea, J., Rosas, A., Lachmann, M., Hannon, G.J., Pääbo, S., 2010. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science* 328, 723–725. <https://doi.org/10.1126/science.1188046>
- Burrill, J., Williams, C.J., Bain, G., Conder, G., Hine, A.L., Misra, R.R., 2007. Tuberculosis: a radiologic review. *Radiographics* 27, 1255–1273. <https://doi.org/10.1148/rg.275065176>
- Canci, A., Minozzi, S., Tarli, S.M.B., 1996. New Evidence of Tuberculous Spondylitis from Neolithic Liguria (Italy). *International Journal of Osteoarchaeology* 6, 497–501. [https://doi.org/10.1002/\(SICI\)1099-1212\(199612\)6:5<497::AID-OA291>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1099-1212(199612)6:5<497::AID-OA291>3.0.CO;2-O)
- Cano, R.J., Rivera-Perez, J., Toranzos, G.A., Santiago-Rodriguez, T.M., Narganes-Storde, Y.M., Chanlatte-Baik, L., García-Roldán, E., Bunkley-Williams, L., Massey, S.E., 2014. Paleomicrobiology: Revealing Fecal Microbiomes of Ancient Indigenous Cultures. *PLOS ONE* 9, e106833. <https://doi.org/10.1371/journal.pone.0106833>
- Chan, J.Z.-M., Sergeant, M.J., Lee, O.Y.-C., Minnikin, D.E., Besra, G.S., Pap, I., Spigelman, M., Donoghue, H.D., Pallen, M.J., 2013. Metagenomic Analysis of Tuberculosis in a Mummy. *New England Journal of Medicine* 369, 289–290. <https://doi.org/10.1056/NEJMc1302295>
- Chen, T., Yu, W.-H., Izard, J., Baranova, O.V., Lakshmanan, A., Dewhirst, F.E., 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database* 2010, baq013. <https://doi.org/10.1093/database/baq013>
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. <https://doi.org/10.4161/fly.19695>
- Cohen, M.N., Armelagos, G.J., 1984. *Paleopathology at the Origins of Agriculture*. University Press of Florida, Gainesville, Florida.
- Coll, F., McNerney, R., Guerra-Assunção, J.A., Glynn, J.R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014. A robust SNP barcode for typing

- Mycobacterium tuberculosis complex strains. *Nature Communications* 5, 4812. <https://doi.org/10.1038/ncomms5812>
- Comas, I., Chakravartti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., Gagneux, S., 2010. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. *Nat Genet* 42, 498–503. <https://doi.org/10.1038/ng.590>
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet* 45, 1176–1182. <https://doi.org/10.1038/ng.2744>
- Consortium, T.H.M.P., 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. <https://doi.org/10.1038/nature11234>
- Correa, A.F., Bailão, A.M., Bastos, I.M.D., Orme, I.M., Soares, C.M.A., Kipnis, A., Santana, J.M., Junqueira-Kipnis, A.P., 2014. The Endothelin System Has a Significant Role in the Pathogenesis and Progression of Mycobacterium tuberculosis Infection. *Infection and Immunity* 82, 5154–5165. <https://doi.org/10.1128/IAI.02304-14>
- Côté, N.M.L., Daligault, J., Pruvost, M., Bennett, E.A., Gorgé, O., Guimaraes, S., Capelli, N., Bailly, M.L., Geigl, E.-M., Grange, T., 2016. A New High-Throughput Approach to Genotype Ancient Human Gastrointestinal Parasites. *PLOS ONE* 11, e0146230. <https://doi.org/10.1371/journal.pone.0146230>
- Crubézy, E., Legal, L., Fabas, G., Dabernat, H., Ludes, B., 2006. Pathogeny of archaic mycobacteria at the emergence of urban life in Egypt (3400 bc). *Infection, Genetics and Evolution* 6, 13–21. <https://doi.org/10.1016/j.meegid.2004.12.003>
- Cui, Y., Yu, C., Yan, Y., Li, D., Li, Yanjun, Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu, L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z., Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, Jing, Yao, S., Rakin, A., Li, Yingrui, Falush, D., Balloux, F., Achtman, M., Song, Y., Wang, Jun, Yang, R., 2013. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *PNAS* 110, 577–582. <https://doi.org/10.1073/pnas.1205750110>
- Dabney, Jesse, Meyer, M., Pääbo, S., 2013. Ancient DNA Damage. *Cold Spring Harb Perspect Biol* 5, a012567. <https://doi.org/10.1101/cshperspect.a012567>
- Dabney, Jesse, Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 110, 15758–15763. <https://doi.org/10.1073/pnas.1314445110>
- De La Fuente, C., Flores, S., Moraga, M., 2013. DNA from human ancient bacteria: A novel source of genetic evidence from archaeological dental calculus. *Archaeometry* 55, 767–778. <https://doi.org/10.1111/j.1475-4754.2012.00707.x>
- Devault, A.M., Golding, G.B., Waglechner, N., Enk, J.M., Kuch, M., Tien, J.H., Shi, M., Fisman, D.N., Dhody, A.N., Forrest, S., Bos, K.I., Earn, D.J.D., Holmes, E.C., Poinar, H.N., 2014. Second-Pandemic Strain of Vibrio cholerae from the Philadelphia Cholera Outbreak of 1849. *New England Journal of Medicine* 370, 334–340. <https://doi.org/10.1056/NEJMoa1308663>
- Devault, A.M., Mortimer, T.D., Kitchen, A., Kiesewetter, H., Enk, J.M., Golding, G.B., Southon, J., Kuch, M., Duggan, A.T., Aylward, W., Gardner, S.N., Allen, J.E., King, A.M., Wright, G., Kuroda, M., Kato, K., Briggs, D.E., Fornaciari, G., Holmes, E.C., Poinar, H.N., Pepperell, C.S., 2017. A molecular portrait of maternal sepsis from Byzantine Troy. *eLife* 6. <https://doi.org/10.7554/eLife.20983>
- Didelot, X., Wilson, D.J., 2015. ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology* 11, e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>

- Dinan, T.G., Stilling, R.M., Stanton, C., Cryan, J.F., 2015. Collective unconscious: How gut microbes shape human behavior. *Journal of Psychiatric Research* 63, 1–9. <https://doi.org/10.1016/j.jpsychires.2015.02.021>
- Dobson, A.P., Carper, E.R., 1996. Infectious Diseases and Human Population History Throughout history the establishment of disease has been a side effect of the growth of civilization. *Bioscience* 46, 115–126.
- Doughty, E.L., Sergeant, M.J., Adetifa, I., Antonio, M., Pallen, M.J., 2014. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ* 2, e585. <https://doi.org/10.7717/peerj.585>
- Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O., Raoult, D., 1998. Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. *PNAS* 95, 12637–12640. <https://doi.org/10.1073/pnas.95.21.12637>
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed Phylogenetics and Dating with Confidence. *PLOS Biology* 4, e88. <https://doi.org/10.1371/journal.pbio.0040088>
- Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. *Trends in Ecology & Evolution* 18, 481–488. [https://doi.org/10.1016/S0169-5347\(03\)00216-7](https://doi.org/10.1016/S0169-5347(03)00216-7)
- Dubos, R., Dubos, J., 1952. *The White Plague: Tuberculosis, Man, and Society*, 3rd ed. Rutgers University Press, New Brunswick, New Jersey.
- Duchêne, S., Duchêne, D., Holmes, E.C., Ho, S.Y.W., 2015. The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data. *Mol Biol Evol* 32, 1895–1906. <https://doi.org/10.1093/molbev/msv056>
- Duchêne, S., Holmes, E.C., Ho, S.Y.W., 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc. B* 281, 20140732. <https://doi.org/10.1098/rspb.2014.0732>
- El-Najjar, M., Al-Shiyab, A., Al-Sarie, I., 1996. Cases of tuberculosis at 'Ain Ghazal, Jordan. *Paléorient* 22, 123–128. <https://doi.org/10.3406/paleo.1996.4639>
- Escapa, I.F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F.E., Lemon, K.P., 2018. New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems* 3, e00187-18. <https://doi.org/10.1128/mSystems.00187-18>
- Faerman, M., Jankauskas, R., 2000. Paleopathological and molecular evidence of human bone tuberculosis in Iron Age Lithuania. *Anthropologischer Anzeiger* 58, 57–62.
- Falkinham, J.O., 2015. Environmental Sources of Nontuberculous Mycobacteria. *Clinics in Chest Medicine* 36, 35–41. <https://doi.org/10.1016/j.ccm.2014.10.003>
- Feldman, M., Harbeck, M., Keller, M., Spyrou, M.A., Rott, A., Trautmann, B., Scholz, H.C., Pääffgen, B., Peters, J., McCormick, M., Bos, K., Herbig, A., Krause, J., 2016. A high-coverage *Yersinia pestis* Genome from a 6th-century Justinianic Plague Victim. *Mol Biol Evol* msw170. <https://doi.org/10.1093/molbev/msw170>
- Filippo, C.D., Cavalieri, D., Paola, M.D., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., Lionetti, P., 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *PNAS* 107, 14691–14696. <https://doi.org/10.1073/pnas.1005963107>
- Fletcher, H.A., 2003. Molecular analysis of *Mycobacterium tuberculosis* DNA from a family of 18th century Hungarians. *Microbiology* 149, 143–151. <https://doi.org/10.1099/mic.0.25961-0>
- Fletcher, H.A., Donoghue, H.D., Michael Taylor, G., van der Zanden, A.G.M., Spigelman, M., 2003. Molecular analysis of *Mycobacterium tuberculosis* DNA from a family of 18th century Hungarians. *Microbiology* 149, 143–151. <https://doi.org/10.1099/mic.0.25961-0>
- Ford, C.B., Lin, P.L., Chase, M.R., Shah, R.R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T.R., Sacchettini, J.C., Lipsitch, M., Flynn, J.L., Fortune, S.M., 2011. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium*

- tuberculosis* during latent infection. *Nature Genetics* 43, 482–486. <https://doi.org/10.1038/ng.811>
- Formicola, V., Milanesi, Q., Scarsini, C., 1987. Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *American Journal of Physical Anthropology* 72, 1–6. <https://doi.org/10.1002/ajpa.1330720102>
- Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire, T.T., Zhu, A., Shao, Y., Pike, L.J., Louie, T., Browne, H.P., Mitchell, A.L., Neville, B.A., Finn, R.D., Lawley, T.D., 2019. A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nature Biotechnology* 37, 186. <https://doi.org/10.1038/s41587-018-0009-7>
- Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., Pääbo, S., 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *PNAS* 110, 2223–2227. <https://doi.org/10.1073/pnas.1221359110>
- Fugassa, M.H., Beltrame, M.O., Sardella, N.H., Civalero, M.T., Aschero, C., 2010. Paleoparasitological results from coprolites dated at the Pleistocene–Holocene transition as source of paleoecological evidence in Patagonia. *Journal of Archaeological Science* 37, 880–884. <https://doi.org/10.1016/j.jas.2009.11.018>
- Gagneux, S., 2018. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology* 16, 202–213. <https://doi.org/10.1038/nrmicro.2018.8>
- Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., Jong, B.C. de, Narayanan, S., Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C., Hilty, M., Hopewell, P.C., Small, P.M., 2006. Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *PNAS* 103, 2869–2873. <https://doi.org/10.1073/pnas.0511240103>
- Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., Meyer, M., 2017. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res* 45, e79–e79. <https://doi.org/10.1093/nar/gkx033>
- Geiss, H.K., Feldhues, R., Niemann, S., Nolte, O., Rieker, R., 2005. Landouzy septicemia (sepsis tuberculosa acutissima) due to *Mycobacterium microti* in an immunocompetent man. *Infection* 33, 393–396. <https://doi.org/10.1007/s15010-005-5075-3>
- Gelabert, P., Sandoval-Velasco, M., Olalde, I., Fregel, R., Rieux, A., Escosa, R., Aranda, C., Paaajmans, K., Mueller, I., Gilbert, M.T.P., Lalueza-Fox, C., 2016. Mitochondrial DNA from the eradicated European *Plasmodium vivax* and *P. falciparum* from 70-year-old slides from the Ebro Delta in Spain. *Proceedings of the National Academy of Sciences* 201611017. <https://doi.org/10.1073/pnas.1611017113>
- Gilbert, M.T.P., Bandelt, H.J., Hofreiter, M., Barnes, I., 2005. Assessing ancient DNA studies. *Trends in Ecology & Evolution* 20, 541–544. <https://doi.org/10.1016/j.tree.2005.07.005>
- Gilbert, M.T.P., Jenkins, D.L., Götherstrom, A., Naveran, N., Sanchez, J.J., Hofreiter, M., Thomsen, P.F., Binladen, J., Higham, T.F.G., Yohe, R.M., Parr, R., Cummings, L.S., Willerslev, E., 2008. DNA from Pre-Clovis Human Coprolites in Oregon, North America. *Science* 320, 786–789. <https://doi.org/10.1126/science.1154116>
- Ginolhac, A., Rasmussen, M., Gilbert, M.T.P., Willerslev, E., Orlando, L., 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* 27, 2153–2155. <https://doi.org/10.1093/bioinformatics/btr347>
- Gonçalves, M.L.C., Araújo, A., Duarte, R., da Silva, J.P., Reinhard, K., Bouchet, F., Ferreira, L.F., 2002. Detection of *Giardia duodenalis* antigen in coprolites using a commercially available enzyme-linked immunosorbent assay. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 96, 640–643. [https://doi.org/10.1016/S0035-9203\(02\)90337-8](https://doi.org/10.1016/S0035-9203(02)90337-8)
- Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., Pääbo, S., 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330. <https://doi.org/10.1038/nature05336>

- Group, T.N.H.W., Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J.A., Bonazzi, V., McEwen, J.E., Wetterstrand, K.A., Deal, C., Baker, C.C., Francesco, V.D., Howcroft, T.K., Karp, R.W., Lunsford, R.D., Wellington, C.R., Belachew, T., Wright, M., Giblin, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J., Little, A.R., Peavy, H., Pontzer, C., Portnoy, M., Sayre, M.H., Starke-Reed, P., Zakhari, S., Read, J., Watson, B., Guyer, M., 2009. The NIH Human Microbiome Project. *Genome Res.* 19, 2317–2323. <https://doi.org/10.1101/gr.096651.109>
- Guellil, M., Kersten, O., Namouchi, A., Bauer, E.L., Derrick, M., Jensen, A.Ø., Stenseth, N.C., Bramanti, B., 2018. Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia. *PNAS* 201807266. <https://doi.org/10.1073/pnas.1807266115>
- Guhl, F., Jaramillo, C., Vallejo, G.A., Yockteng, R., Cardenas-Arroyo, F., Fornaciari, G., Arriaza, B., Aufderheide, A.C., 1999. Isolation of *Trypanosoma cruzi* DNA in 4,000-year-old mummified human tissue from northern Chile. *American Journal of Physical Anthropology* 108, 401–407.
- Gutacker, M.M., Smoot, J.C., Migliaccio, C.A.L., Ricklefs, S.M., Hua, S., Cousins, D.V., Graviss, E.A., Shashkina, E., Kreiswirth, B.N., Musser, J.M., 2002. Genome-Wide Analysis of Synonymous Single Nucleotide Polymorphisms in *Mycobacterium tuberculosis* Complex Organisms: Resolution of Genetic Relationships Among Closely Related Microbial Strains. *Genetics* 162, 1533–1543.
- Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., Supply, P., Vincent, V., 2005. Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. *PLOS Pathogens* 1, e5. <https://doi.org/10.1371/journal.ppat.0010005>
- Han, E.-T., Guk, S.-M., Kim, J.-L., Jeong, H.-J., Kim, S.-N., Chai, J.-Y., 2003. Detection of parasite eggs from archaeological excavations in the Republic of Korea. *Mem. Inst. Oswaldo Cruz* 98, 123–126. <https://doi.org/10.1590/S0074-02762003000900018>
- He, F., Hu, X.-S., 2005. Hubbell's fundamental biodiversity parameter and the Simpson diversity index. *Ecology Letters* 8, 386–390. <https://doi.org/10.1111/j.1461-0248.2005.00729.x>
- Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W., Gagneux, S., 2008. High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography. *PLOS Biology* 6, e311. <https://doi.org/10.1371/journal.pbio.0060311>
- Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y.-C., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., Bar-Gal, G.K., Spigelman, M., 2008. Detection and Molecular Characterization of 9000-Year-Old *Mycobacterium tuberculosis* from a Neolithic Settlement in the Eastern Mediterranean. *PLoS ONE* 3, e3426. <https://doi.org/10.1371/journal.pone.0003426>
- Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., Wilson, A.C., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312.
- Ho, S.Y.W., Larson, G., 2006. Molecular clocks: when times are a-changin'. *Trends in Genetics* 22, 79–83. <https://doi.org/10.1016/j.tig.2005.11.006>
- Ho, S.Y.W., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Mol Biol Evol* 22, 1561–1568. <https://doi.org/10.1093/molbev/msi145>
- Ho, S.Y.W., Shapiro, B., Phillips, M.J., Cooper, A., Drummond, A.J., 2007. Evidence for Time Dependency of Molecular Rate Estimates. *Syst Biol* 56, 515–522. <https://doi.org/10.1080/10635150701435401>
- Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Benjamin Gordon, D., Brizuela, L., Richard McCombie, W., Hannon, G.J., 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat. Protocols* 4, 960–974. <https://doi.org/10.1038/nprot.2009.68>
- Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. von, Pääbo, S., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in

- ancient DNA. *Nucleic Acids Res* 29, 4793–4799.
<https://doi.org/10.1093/nar/29.23.4793>
- Hofreiter, M., Poinar, H.N., Spaulding, W.G., Bauer, K., Martin, P.S., Possnert, G., Pääbo, S., 2000. A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* 9, 1975–1984. <https://doi.org/10.1046/j.1365-294X.2000.01106.x>
- Houben, R.M.G.J., Dodd, P.J., 2016. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Medicine* 13, e1002152.
<https://doi.org/10.1371/journal.pmed.1002152>
- Hruska, K., Kaevska, M., 2013. Mycobacteria in water, soil, plants and air: a review. *Veterinari Medicina* 57, 623–679. <https://doi.org/10.17221/6558-VETMED>
- Huebler, R., Key, F.M., Warinner, C., Bos, K.I., Krause, J., Herbig, A., 2019. HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *bioRxiv* 534198. <https://doi.org/10.1101/534198>
- Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* 12, e1004957.
<https://doi.org/10.1371/journal.pcbi.1004957>
- Hyde, E.R., Haarmann, D.P., Lynne, A.M., Bucheli, S.R., Petrosino, J.F., 2013. The Living Dead: Bacterial Community Structure of a Cadaver at the Onset and End of the Bloat Stage of Decomposition. *PLOS ONE* 8, e77733.
<https://doi.org/10.1371/journal.pone.0077733>
- Iñiguez, A.M., Reinhard, K., Carvalho Gonçalves, M.L., Ferreira, L.F., Araújo, A., Paulo Vicente, A.C., 2006. SL1 RNA gene recovery from *Enterobius vermicularis* ancient DNA in pre-Columbian human coprolites. *International Journal for Parasitology* 36, 1419–1425. <https://doi.org/10.1016/j.ijpara.2006.07.005>
- Iñiguez, A.M., Reinhard, K.J., Araújo, A., Ferreira, L.F., Vicente, A.C.P., 2003. *Enterobius vermicularis*: ancient DNA from north and south American human coprolites. *Memórias do Instituto Oswaldo Cruz* 98, 67–69. <https://doi.org/10.1590/S0074-02762003000900013>
- International Helminth Genomes Consortium, 2019. Comparative genomics of the major parasitic worms. *Nature Genetics* 51, 163. <https://doi.org/10.1038/s41588-018-0262-1>
- Irwin, K.K., Laurent, S., Matuszewski, S., Vuilleumier, S., Ormond, L., Shim, H., Bank, C., Jensen, J.D., 2016. On the importance of skewed offspring distributions and background selection in virus population genetics. *Heredity* 117, 393–399.
<https://doi.org/10.1038/hdy.2016.58>
- Jankute, M., Nataraj, V., Lee, O.Y.-C., Wu, H.H.T., Ridell, M., Garton, N.J., Barer, M.R., Minnikin, D.E., Bhatt, A., Besra, G.S., 2017. The role of hydrophobicity in tuberculosis evolution and pathogenicity. *Scientific Reports* 7, 1315.
<https://doi.org/10.1038/s41598-017-01501-0>
- Jenkins, D.L., Davis, L.G., Stafford, T.W., Campos, P.F., Hockett, B., Jones, G.T., Cummings, L.S., Yost, C., Connolly, T.J., Yohe, R.M., Gibbons, S.C., Raghavan, M., Rasmussen, M., Paijmans, J.L.A., Hofreiter, M., Kemp, B.M., Barta, J.L., Monroe, C., Gilbert, M.T.P., Willerslev, E., 2012. Clovis Age Western Stemmed Projectile Points and Human Coprolites at the Paisley Caves. *Science* 337, 223–228.
<https://doi.org/10.1126/science.1218443>
- Kapopoulou, A., Lew, J.M., Cole, S.T., 2011. The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis, SPECIAL ISSUE: Including a Special section: Inhaled Therapies for TB* 91, 8–13.
<https://doi.org/10.1016/j.tube.2010.09.006>
- Kay, G.L., Sergeant, M.J., Giuffra, V., Bandiera, P., Milanese, M., Bramanti, B., Bianucci, R., Pallen, M.J., 2014. Recovery of a Medieval *Brucella melitensis* Genome Using Shotgun Metagenomics. *mBio* 5, e01337-14. <https://doi.org/10.1128/mBio.01337-14>
- Kay, G.L., Sergeant, M.J., Zhou, Z., Chan, J.Z.-M., Millard, A., Quick, J., Szikossy, I., Pap, I., Spigelman, M., Loman, N.J., Achtman, M., Donoghue, H.D., Pallen, M.J., 2015. Eighteenth-century genomes show that mixed infections were common at time of

- peak tuberculosis in Europe. *Nature Communications* 6, 6717. <https://doi.org/10.1038/ncomms7717>
- Keane, T., Creevey, C., Pentony, M., Naughton, T.J., McInerney, J., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology* 6, 29–46.
- Kebschull, M., Demmer, R.T., Papapanou, P.N., 2010. “Gum Bug, Leave My Heart Alone!”—Epidemiologic and Mechanistic Evidence Linking Periodontal Infections and Atherosclerosis. *J Dent Res* 89, 879–902. <https://doi.org/10.1177/0022034510375281>
- Keller, M., Spyrou, M.A., Scheib, C.L., Kröpelin, A., Haas-Gebhard, B., Pääffgen, B., Haberstroh, J., Lacomba, A.R. i, Raynaud, C., Cessford, C., Stadler, P., Nägele, K., Neumann, G.U., Bates, J.S., Trautmann, B., Inskip, S., Peters, J., Robb, J.E., Kivisild, T., McCormick, M., Bos, K.I., Harbeck, M., Herbig, A., Krause, J., 2018. Ancient *Yersinia pestis* genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). *bioRxiv* 481226. <https://doi.org/10.1101/481226>
- Key, F.M., Posth, C., Krause, J., Herbig, A., Bos, K.I., 2017. Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. *Trends in Genetics* 33, 508–520. <https://doi.org/10.1016/j.tig.2017.05.005>
- Kim, Y.-S., Lee, I.S., Oh, C.S., Kim, M.J., Cha, S.C., Shin, D.H., 2016. Calcified Pulmonary Nodules Identified in a 350-Year-Old-Joseon Mummy: the First Report on Ancient Pulmonary Tuberculosis from Archaeologically Obtained Pre-modern Korean Samples. *Journal of Korean Medical Science* 31, 147–151. <https://doi.org/10.3346/jkms.2016.31.1.147>
- Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* 40, e3–e3. <https://doi.org/10.1093/nar/gkr771>
- Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman, F.D., Knight, R., Kelley, S.T., 2011. Bayesian community-wide culture-independent microbial source tracking. *Nat Meth* 8, 761–763. <https://doi.org/10.1038/nmeth.1650>
- Köhler, K., Pálfi, G., Molnár, E., Zalai-Gaál, I., Oszrás, A., Bánffy, E., Kirinó, K., Kiss, K.K., Mende, B.G., 2014. A Late Neolithic Case of Pott’s Disease from Hungary. *International Journal of Osteoarchaeology* 24, 697–703. <https://doi.org/10.1002/oa.2254>
- Koren, O., Spor, A., Felin, J., Fåk, F., Stombaugh, J., Tremaroli, V., Behre, C.J., Knight, R., Fagerberg, B., Ley, R.E., Bäckhed, F., 2011. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *PNAS* 108, 4592–4598. <https://doi.org/10.1073/pnas.1011383107>
- Kostyukevich, Y., Bugrova, A., Chagovets, V., Brzhozovskiy, A., Indeykina, M., Vanyushkina, A., Zherebker, A., Mitina, A., Kononikhin, A., Popov, I., Khaitovich, P., Nikolaev, E., 2018. Proteomic and lipidomic analysis of mammoth bone by high-resolution tandem mass spectrometry coupled with liquid chromatography. *Eur J Mass Spectrom (Chichester)* 24, 411–419. <https://doi.org/10.1177/1469066718813728>
- Krause, J., Briggs, A.W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., Pääbo, S., 2010. A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Current Biology* 20, 231–236. <https://doi.org/10.1016/j.cub.2009.11.068>
- Krause-Kyora, B., Susat, J., Key, F.M., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D., 2018. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B 15.
- Kumar, S., Stecher, G., Peterson, D., Tamura, K., 2012. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 28, 2685–2686. <https://doi.org/10.1093/bioinformatics/bts507>
- Le Bailly, M., Gonçalves, M.L., Harter-Lailheugue, S., Prodéo, F., Araujo, A., Bouchet, F., 2008. New finding of *Giardia intestinalis* (Eukaryote, Metamonad) in Old World archaeological site using immunofluorescence and enzyme-linked immunosorbent

- assays. *Mem. Inst. Oswaldo Cruz* 103, 298–300. <https://doi.org/10.1590/S0074-02762008005000018>
- Lee, O.Y.-C., Wu, H.H.T., Besra, G.S., Rothschild, B.M., Spigelman, M., Hershkovitz, I., Bar-Gal, G.K., Donoghue, H.D., Minnikin, D.E., 2015. Lipid biomarkers provide evolutionary signposts for the oldest known cases of tuberculosis. *Tuberculosis, Supplement issue: Tuberculosis in Evolution* 95, S127–S132. <https://doi.org/10.1016/j.tube.2015.02.013>
- Lee, O.Y.-C., Wu, H.H.T., Donoghue, H.D., Spigelman, M., Greenblatt, C.L., Bull, I.D., Rothschild, B.M., Martin, L.D., Minnikin, D.E., Besra, G.S., 2012. Mycobacterium tuberculosis Complex Lipid Virulence Factors Preserved in the 17,000-Year-Old Skeleton of an Extinct Bison, *Bison antiquus*. *PLOS ONE* 7, e41923. <https://doi.org/10.1371/journal.pone.0041923>
- Leles, D., Araújo, A., Ferreira, L.F., Vicente, A.C.P., Iñiguez, A.M., 2008. Molecular paleoparasitological diagnosis of *Ascaris* sp. from coprolites: new scenery of ascariasis in pre-Colombian South America times. *Memórias do Instituto Oswaldo Cruz* 103, 106–108. <https://doi.org/10.1590/S0074-02762008005000004>
- Leung, A.N., Müller, N.L., Pineda, P.R., FitzGerald, J.M., 1992. Primary tuberculosis in childhood: radiographic manifestations. *Radiology* 182, 87–91. <https://doi.org/10.1148/radiology.182.1.1727316>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.
- Liu, Q., Ma, A., Wei, L., Pang, Y., Wu, B., Luo, T., Zhou, Y., Zheng, H.-X., Jiang, Q., Gan, M., Zuo, T., Liu, M., Yang, C., Jin, L., Comas, I., Gagneux, S., Zhao, Y., Pepperell, C.S., Gao, Q., 2018. China's tuberculosis epidemic stems from historical expansion of four strains of *Mycobacterium tuberculosis*. *Nature Ecology & Evolution* 1. <https://doi.org/10.1038/s41559-018-0680-6>
- Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., Stretton, S., Brock, F., Higham, T., Park, Y., Hofreiter, M., Bradley, D.G., Bhak, J., Pinhasi, R., Manica, A., 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* 350, 820–822. <https://doi.org/10.1126/science.aad2879>
- Loosdrecht, M. van de, Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., Hajraoui, M.A.E., Amzazi, S., Hublin, J.-J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W., Jeong, C., Krause, J., 2018. Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* 360, 548–552. <https://doi.org/10.1126/science.aar8380>
- Loreille, O., Roumat, E., Verneau, O., Bouchet, F., Hänni, C., 2001. Ancient DNA from *Ascaris*: extraction amplification and sequences from eggs collected in coprolites. *International Journal for Parasitology* 31, 1101–1106. [https://doi.org/10.1016/S0020-7519\(01\)00214-4](https://doi.org/10.1016/S0020-7519(01)00214-4)
- Lugli, G.A., Milani, C., Mancabelli, L., Turrone, F., Ferrario, C., Duranti, S., van Sinderen, D., Ventura, M., 2017. Ancient bacteria of the Ötzi's microbiome: a genomic tale from the Copper Age. *Microbiome* 5. <https://doi.org/10.1186/s40168-016-0221-y>
- Lunardini, A., Costantini, L., Biasini, L.C., Caramella, D., Fornaciari, G., 2012. Evidence of Congenital Syphilis and Tuberculosis in a XIX Century Mummy (Perugia, Italy) 3.
- Maicher, C., Hoffmann, A., Côté, N.M., Palomo Pérez, A., Saña Seguí, M., Le Bailly, M., 2017. Paleoparasitological investigations on the Neolithic lakeside settlement of La Draga (Lake Banyoles, Spain). *The Holocene* 0959683617702236.
- Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M.R., Hallows, J.L., Kusebauch, U., Vigl, E.E., Malfertheiner, P., Megraud, F., O'Sullivan, N., Cipollini, G., Coia, V., Samadelli, M., Engstrand, L., Linz, B., Moritz, R.L., Grimm, R., Krause, J., Nebel, A., Moodley, Y., Rattei, T., Zink, A., 2016. The 5300-year-old *Helicobacter*

- pylori genome of the Iceman. *Science* 351, 162–165.
<https://doi.org/10.1126/science.aad2545>
- Mann, A., 2018. Novel Techniques for the Description and Interpretation of Microbial Communities in the Modern Human Gut and Ancient Human Oral Microbiome.
- Mann, A.E., Sabin, S., Ziesemer, K., Vågane, Å.J., Schroeder, H., Ozga, A.T., Sankaranarayanan, K., Hofman, C.A., Yates, J.A.F., Salazar-García, D.C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G.R., Stone, A.C., Lewis, C.M., Krause, J., Hofman, C., Bos, K.I., Warinner, C., 2018. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific Reports* 8, 9822. <https://doi.org/10.1038/s41598-018-28091-9>
- Marciniak, S., Prowse, T.L., Herring, D.A., Klunk, J., Kuch, M., Duggan, A.T., Bondioli, L., Holmes, E.C., Poinar, H.N., 2016. Plasmodium falciparum malaria in 1st–2nd century CE southern Italy. *Current Biology* 26, R1220–R1222.
<https://doi.org/10.1016/j.cub.2016.10.016>
- Mardis, E.R., 2008. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* 9, 387–402.
<https://doi.org/10.1146/annurev.genom.9.081307.164359>
- Maricic, T., Whitten, M., Pääbo, S., 2010. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLOS ONE* 5, e14004.
<https://doi.org/10.1371/journal.pone.0014004>
- Masson, M., Molnár, E., Donoghue, H.D., Besra, G.S., Minnikin, D.E., Wu, H.H.T., Lee, O.Y.-C., Bull, I.D., Pálfi, G., 2013. Osteological and Biomolecular Evidence of a 7000-Year-Old Case of Hypertrophic Pulmonary Osteopathy Secondary to Tuberculosis from Neolithic Hungary. *PLOS ONE* 8, e78252.
- Mathuria, J.P., Sharma, P., Prakash, P., Samaria, J.K., Katoch, V.M., Anupurba, S., 2008. Role of spoligotyping and IS6110-RFLP in assessing genetic diversity of Mycobacterium tuberculosis in India. *Infection, Genetics and Evolution* 8, 346–351.
<https://doi.org/10.1016/j.meegid.2008.02.005>
- Mazade, M., Evans, E., Starke, J., Correa, A., 2001. Congenital tuberculosis presenting as sepsis syndrome: case report and review of the literature. *The Pediatric Infectious Disease Journal* 20, 439–442.
- Menon, R., Ramanan, V., Korolev, K.S., 2018. Interactions between species introduce spurious associations in microbiome studies. *PLOS Computational Biology* 14, e1005939. <https://doi.org/10.1371/journal.pcbi.1005939>
- Meyer, M., Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* 2010, pdb.prot5448-pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Michel, P., Barbier, C., Loubière, Y., Hayon, J.-H., Ricôme, J.-L., 2002. Three cases of septic shock due to tuberculosis without HIV pathology. *Intensive Care Med* 28, 1827–1828.
<https://doi.org/10.1007/s00134-002-1526-9>
- Mitchell, P.D., 2013. The origins of human parasites: Exploring the evidence for endoparasitism throughout human evolution. *International Journal of Paleopathology* 3, 191–198. <https://doi.org/10.1016/j.ijpp.2013.08.003>
- Mitchell, P.D., 2011. Retrospective diagnosis and the use of historical texts for investigating disease in the past. *International Journal of Paleopathology* 1, 81–88.
<https://doi.org/10.1016/j.ijpp.2011.04.002>
- Mitchell, P.D., Anastasiou, E., Syon, D., 2011. Human intestinal parasites in crusader Acre: Evidence for migration with disease in the medieval period. *International Journal of Paleopathology* 1, 132–137. <https://doi.org/10.1016/j.ijpp.2011.10.005>
- Mitchell, P.D., Stern, E., Tepper, Y., 2008. Dysentery in the crusader kingdom of Jerusalem: an ELISA analysis of two medieval latrines in the City of Acre (Israel). *Journal of Archaeological Science* 35, 1849–1853. <https://doi.org/10.1016/j.jas.2007.11.017>
- Möller, S., Plessis, L. du, Stadler, T., 2018. Impact of the tree prior on estimating clock rates during epidemic outbreaks. *PNAS* 201713314.
<https://doi.org/10.1073/pnas.1713314115>

- Mostowy, S., Cousins, D., Brinkman, J., Aranaz, A., Behr, M.A., 2002. Genomic Deletions Suggest a Phylogeny for the *Mycobacterium tuberculosis* Complex. *J Infect Dis* 186, 74–80. <https://doi.org/10.1086/341068>
- Mühlemann, B., Jones, T.C., Damgaard, P. de B., Allentoft, M.E., Shevna, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., Tashbaeva, K., Merz, V., Lau, N., Smrčka, V., Voyakin, D., Kitov, E., Epimakhov, A., Pokutta, D., Vicze, M., Price, T.D., Moiseyev, V., Hansen, A.J., Orlando, L., Rasmussen, S., Sikora, M., Vinner, L., Osterhaus, A.D.M.E., Smith, D.J., Glebe, D., Fouchier, R.A.M., Drosten, C., Sjögren, K.-G., Kristiansen, K., Willerslev, E., 2018. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* 557, 418. <https://doi.org/10.1038/s41586-018-0097-z>
- Müller, R., Roberts, C.A., Brown, T.A., 2015. Complications in the study of ancient tuberculosis: non-specificity of IS6110 PCRs. *STAR: Science & Technology of Archaeological Research* 1, 1–8. <https://doi.org/10.1179/2054892314Y.0000000002>
- Mullis, K.B., Faloona, F.A., 1987. [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, in: *Methods in Enzymology, Recombinant DNA Part F*. Academic Press, pp. 335–350. [https://doi.org/10.1016/0076-6879\(87\)55023-6](https://doi.org/10.1016/0076-6879(87)55023-6)
- Neukamm, J., Peltzer, A., 2018. Integrative-Transcriptomics/DamageProfiler.
- Nielsen, H.B., Almeida, M., Juncker, Agnieszka Sierakowska, Rasmussen, S., Li, J., Sunagawa, S., Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., Quintanilha dos Santos, M.B., Blom, N., Borruel, N., Burgdorf, K.S., Boumezbeur, F., Casellas, F., Doré, J., Dworzynski, P., Guarner, F., Hansen, T., Hildebrand, F., Kaas, R.S., Kennedy, S., Kristiansen, K., Kultima, J.R., Léonard, P., Levenez, F., Lund, O., Moumen, B., Le Paslier, D., Pons, N., Pedersen, O., Prifti, E., Qin, J., Raes, J., Sørensen, S., Tap, J., Tims, S., Ussery, D.W., Yamada, T., MetaHIT Consortium, Nielsen, H.B., Almeida, M., Juncker, Agnieszka S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D.R., Gautier, L., Pedersen, A.G., Le Chatelier, E., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., Quintanilha dos Santos, M.B., Blom, N., Borruel, N., Burgdorf, K.S., Boumezbeur, F., Casellas, F., Doré, J., Dworzynski, P., Guarner, F., Hansen, T., Hildebrand, F., Kaas, R.S., Kennedy, S., Kristiansen, K., Kultima, J.R., Leonard, P., Levenez, F., Lund, O., Moumen, B., Le Paslier, D., Pons, N., Pedersen, O., Prifti, E., Qin, J., Raes, J., Sørensen, S., Tap, J., Tims, S., Ussery, D.W., Yamada, T., Renault, P., Sicheritz-Ponten, T., Bork, P., Wang, J., Brunak, S., Ehrlich, S.D., Jamet, A., Mérieux, A., Cultrone, A., Torrejon, A., Quinquis, B., Brechot, C., Delorme, C., M'Rini, C., de Vos, W.M., Maguin, E., Varela, E., Guedon, E., Gwen, F., Haimet, F., Artiguenave, F., Vandemeulebrouck, G., Denariáz, G., Khaci, G., Blottière, H., Knol, J., Weissenbach, J., van Hylckama Vlieg, J.E.T., Torben, J., Parkhill, J., Turner, K., van de Guchte, M., Antolin, M., Rescigno, M., Kleerebezem, M., Derrien, M., Galleron, N., Sanchez, N., Grarup, N., Veiga, P., Oozeer, R., Dervyn, R., Layec, S., Bruls, T., Winogradski, Y., Erwin G, Z., Renault, P., Sicheritz-Ponten, T., Bork, P., Wang, J., Brunak, S., Ehrlich, S.D., 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology* 32, 822–828. <https://doi.org/10.1038/nbt.2939>
- Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P.M., Spicer, P., Lawson, P., Marin-Reyes, L., Trujillo-Villarreal, O., Foster, M., Guija-Poma, E., Troncoso-Corzo, L., Warinner, C., Ozga, A.T., Lewis, C.M., 2015. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Communications* 6, 6505. <https://doi.org/10.1038/ncomms7505>
- Ochman, H., Elwyn, S., Moran, N.A., 1999. Calibrating bacterial evolution. *PNAS* 96, 12638–12643. <https://doi.org/10.1073/pnas.96.22.12638>
- Oh, C.S., Seo, M., Chai, J.Y., Lee, S.J., Kim, M.J., Park, J.B., Shin, D.H., 2010a. Amplification and sequencing of *Trichuris trichiura* ancient DNA extracted from

- archaeological sediments. *Journal of Archaeological Science* 37, 1269–1273. <https://doi.org/10.1016/j.jas.2009.12.029>
- Oh, C.S., Seo, M., Lim, N.J., Lee, S.J., Lee, E.-J., Lee, S.D., Shin, D.H., 2010b. Paleoparasitological report on *Ascaris* aDNA from an ancient East Asian sample. *Memórias do Instituto Oswaldo Cruz* 105, 225–228. <https://doi.org/10.1590/S0074-02762010000200020>
- Oh, J., Byrd, A.L., Park, M., Kong, H.H., Segre, J.A., 2016. Temporal Stability of the Human Skin Microbiome. *Cell* 165, 854–866. <https://doi.org/10.1016/j.cell.2016.04.008>
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. *vegan: Community Ecology Package*.
- Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 385. <https://doi.org/10.1186/1471-2105-12-385>
- O'Neill, M.B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2019. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *Mol Ecol* 15120. <https://doi.org/10.1111/mec.15120>
- O'Neill, M.B., Shockey, A.C., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2018. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. *bioRxiv* 210161. <https://doi.org/10.1101/210161>
- Ozga, A.T., Nieves-Colón, M.A., Honap, T.P., Sankaranarayanan, K., Hofman, C.A., Milner, G.R., Lewis, C.M., Stone, A.C., Warinner, C., 2016. Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *Am. J. Phys. Anthropol.* 160, 220–228. <https://doi.org/10.1002/ajpa.22960>
- Pääbo, S., 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *PNAS* 86, 1939–1943. <https://doi.org/10.1073/pnas.86.6.1939>
- Pääbo, S., 1984. Über den Nachweis von DNA in altägyptischen Mumien. *Das Altertum* 30, 213–218.
- Park, J.-K., Kim, K.-H., Kang, S., Jeon, H.K., Kim, J.-H., Littlewood, D.T.J., Eom, K.S., 2007. Characterization of the mitochondrial genome of *Diphyllobothrium latum* (Cestoda: Pseudophyllidea) - implications for the phylogeny of eucestodes. *Parasitology* 134, 749–759. <https://doi.org/10.1017/S003118200600206X>
- Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M.C., Rice, B.L., DuLong, C., Morgan, X.C., Golden, C.D., Quince, C., Huttenhower, C., Segata, N., 2019. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* 176, 649-662.e20. <https://doi.org/10.1016/j.cell.2019.01.001>
- Pearce-Duvet, J.M.C., 2006. The origin of human pathogens: evaluating the role of agriculture and domestic animals in the evolution of human disease. *Biological Reviews* 81, 369. <https://doi.org/10.1017/S1464793106007020>
- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., Nieselt, K., 2016. EAGER: efficient ancient genome reconstruction. *Genome Biology* 17, 60. <https://doi.org/10.1186/s13059-016-0918-z>
- Pène, F., Papo, T., Brudy-Gulphe, L., Cariou, A., Piette, J.C., Vinsonneau, C., 2001. Septic shock and thrombotic microangiopathy due to *Mycobacterium tuberculosis* in a nonimmunocompromised patient. *Arch. Intern. Med.* 161, 1347–1348.
- Pepperell, C.S., Casto, A.M., Kitchen, A., Granka, J.M., Comejo, O.E., Holmes, E.C., Birren, B., Galagan, J., Feldman, M.W., 2013. The Role of Selection in Shaping Diversity of Natural *M. tuberculosis* Populations. *PLOS Pathogens* 9, e1003543. <https://doi.org/10.1371/journal.ppat.1003543>
- Perry, G.H., 2014. Parasites and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews* 23, 218–228. <https://doi.org/10.1002/evan.21427>

- Poinar, H., Kuch, M., McDonald, G., Martin, P., Pääbo, S., 2003. Nuclear Gene Sequences from a Late Pleistocene Sloth Coprolite. *Current Biology* 13, 1150–1152. [https://doi.org/10.1016/S0960-9822\(03\)00450-0](https://doi.org/10.1016/S0960-9822(03)00450-0)
- Poinar, H.N., Hofreiter, M., Spaulding, W.G., Martin, P.S., Stankiewicz, B.A., Bland, H., Evershed, R.P., Possnert, G., Pääbo, S., 1998. Molecular Coproscopy: Dung and Diet of the Extinct Ground Sloth *Nothrotheriops shastensis*. *Science* 281, 402–406. <https://doi.org/10.1126/science.281.5375.402>
- Poinar, H.N., Kuch, M., Sobolik, K.D., Barnes, I., Stankiewicz, A.B., Kuder, T., Spaulding, W.G., Bryant, V.M., Cooper, A., Paabo, S., 2001. A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences* 98, 4317–4322. <https://doi.org/10.1073/pnas.061014798>
- Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R.D.E., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A., Rampp, M., Miller, W., Schuster, S.C., 2006. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science* 311, 392–394. <https://doi.org/10.1126/science.1123360>
- Posth, C., Nägele, K., Colleran, H., Valentin, F., Bedford, S., Kami, K.W., Shing, R., Buckley, H., Kinaston, R., Walworth, M., Clark, G.R., Reepmeyer, C., Flexner, J., Maric, T., Moser, J., Gresky, J., Kiko, L., Robson, K.J., Auckland, K., Oppenheimer, S.J., Hill, A.V.S., Mentzer, A.J., Zech, J., Petchey, F., Roberts, P., Jeong, C., Gray, R.D., Krause, J., Powell, A., 2018. Language continuity despite population replacement in Remote Oceania. *Nature Ecology & Evolution* 1. <https://doi.org/10.1038/s41559-018-0498-2>
- Preus, H.R., Marvik, O.J., Selvig, K.A., Bennike, P., 2011. Ancient bacterial DNA (aDNA) in dental calculus from archaeological human remains. *Journal of Archaeological Science* 38, 1827–1831. <https://doi.org/10.1016/j.jas.2011.03.020>
- Purdy, K.J., Embley, T.M., Takii, S., Nedwell, D.B., 1996. Rapid Extraction of DNA and rRNA from Sediments by a Novel Hydroxyapatite Spin-Column Method. *Appl. Environ. Microbiol.* 62, 3905–3907.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, Shaochuan, Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Paslier, D.L., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, Shengting, Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, Songgang, Qin, N., Yang, H., Wang, Jian, Brunak, S., Doré, J., Guamer, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Consortium, M., Antolin, M., Artiguenave, F., Blottiere, H., Borruel, N., Bruls, T., Casellas, F., Chervaux, C., Cultrone, A., Delorme, C., Denariáz, G., Dervyn, R., Forte, M., Friss, C., Guchte, M. van de, Guedon, E., Haimet, F., Jamet, A., Juste, C., Kaci, G., Kleerebezem, M., Knol, J., Kristensen, M., Layec, S., Roux, K.L., Leclerc, M., Maguin, E., Minardi, R.M., Oozeer, R., Rescigno, M., Sanchez, N., Tims, S., Torrejon, T., Varela, E., Vos, W. de, Winogradsky, Y., Zoetendal, E., Bork, P., Ehrlich, S.D., Wang, Jun, 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59. <https://doi.org/10.1038/nature08821>
- Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2. <https://doi.org/10.1093/ve/vew007>
- Rambaut, A., Suchard, M., Xie, W., Drummond, A.J., 2014. Tracer. Institute of Evolutionary Biology, University of Edinburgh.
- Rampelli, S., Schnorr, S.L., Consolandi, C., Turrioni, S., Severgnini, M., Peano, C., Brigidi, P., Crittenden, A.N., Henry, A.G., Candela, M., 2015. Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Current Biology* 25, 1682–1693. <https://doi.org/10.1016/j.cub.2015.04.055>
- Ramsden, C., Holmes, E.C., Charleston, M.A., 2009. Hantavirus Evolution in Relation to Its Rodent and Insectivore Hosts: No Evidence for Codivergence. *Mol Biol Evol* 26, 143–153. <https://doi.org/10.1093/molbev/msn234>

- Raoult, D., Aboudharam, G., Crubézy, E., Larrouy, G., Ludes, B., Drancourt, M., 2000. Molecular identification by “suicide PCR” of *Yersinia pestis* as the agent of Medieval Black Death. *PNAS* 97, 12800–12803. <https://doi.org/10.1073/pnas.220225197>
- Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., Rasmussen, S., 2019. Emergence and Spread of Basal Lineages of *Yersinia pestis* during the Neolithic Decline. *Cell* 176, 295–305.e10. <https://doi.org/10.1016/j.cell.2018.11.005>
- Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., Nielsen, H.B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M.V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul, L., Yepiskoposyan, L., Sicheritz-Pontén, T., Foley, R.A., Lahr, M.M., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* 163, 571–582. <https://doi.org/10.1016/j.cell.2015.10.009>
- Reinhard, K.J., Hevly, R.H., Anderson, G.A., 1987. Helminth Remains from Prehistoric Indian Coprolites on the Colorado Plateau. *The Journal of Parasitology* 73, 630. <https://doi.org/10.2307/3282147>
- Renaud, G., Slon, V., Duggan, A.T., Kelso, J., 2015. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology* 16, 1–18. <https://doi.org/10.1186/s13059-015-0776-0>
- Richards, M.B., Sykes, B.C., Hedges, R.E.M., 1995. Authenticating DNA Extracted From Ancient Skeletal Remains. *Journal of Archaeological Science* 22, 291–299. <https://doi.org/10.1006/jasc.1995.0031>
- Roberts, C., Ingham, S., 2008. Using ancient DNA analysis in palaeopathology: a critical analysis of published papers, with recommendations for future work. *International Journal of Osteoarchaeology* 18, 600–613. <https://doi.org/10.1002/oa.966>
- Roberts, C.A., Buikstra, J.E., 2003. *The Bioarchaeology of Tuberculosis: A Global View on a Reemerging Disease*. University Press of Florida, Gainesville, Florida.
- Ross, Z.P., Klunk, J., Fornaciari, G., Giuffra, V., Duchêne, S., Duggan, A.T., Poinar, D., Douglas, M.W., Eden, J.-S., Holmes, E.C., Poinar, H.N., 2018. The paradox of HBV evolution as revealed from a 16th century mummy. *PLOS Pathogens* 14, e1006750. <https://doi.org/10.1371/journal.ppat.1006750>
- Rothschild, B.M., Martin, L., Lev, G., Bercovier, H., Bar-Gal, G.K., Greenblatt, C.L., Donoghue, H.D., Spigelman, M., Brittain, D., 2001. *Mycobacterium tuberculosis* Complex DNA from an Extinct Bison Dated 17,000 Years before the Present. *Clinical Infectious Diseases* 305–311.
- Rugg, J., 2000. Defining the place of burial: What makes a cemetery a cemetery? *Mortality* 5, 259–275. <https://doi.org/10.1080/713686011>
- Russell, D.G., 2007. Who puts the tubercle in tuberculosis? *Nat Rev Micro* 5, 39–47. <https://doi.org/10.1038/nrmicro1538>
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisine, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., Boucher, C.A., 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 415, 497–502. <https://doi.org/10.1038/415497a>
- Salo, W.L., Aufderheide, A.C., Buikstra, J., Holcomb, T.A., 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. *PNAS* 91, 2091–2094. <https://doi.org/10.1073/pnas.91.6.2091>
- Sankaranarayanan, K., Ozga, A.T., Warinner, C., Tito, R.Y., Obregon-Tito, A.J., Xu, J., Gaffney, P.M., Jervis, L.L., Cox, D., Stephens, L., Foster, M., Tallbull, G., Spicer, P., Lewis, C.M., 2015. Gut Microbiome Diversity among Cheyenne and Arapaho Individuals from Western Oklahoma. *Current Biology* 25, 3161–3169. <https://doi.org/10.1016/j.cub.2015.10.060>

- Santiago-Rodriguez, T.M., Fornaciari, G., Luciani, S., Dowd, S.E., Toranzos, G.A., Marota, I., Cano, R.J., 2015. Gut Microbiome of an 11th Century A.D. Pre-Columbian Andean Mummy. *PLOS ONE* 10, e0138135. <https://doi.org/10.1371/journal.pone.0138135>
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., Pääbo, S., 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS ONE* 7, e34131. <https://doi.org/10.1371/journal.pone.0034131>
- Schnorr, S.L., Sankaranarayanan, K., Lewis Jr., C.M., Warinner, C., 2016. Insights into human evolution from ancient and contemporary microbiome studies. *Current Opinion in Genetics & Development, Genetics of human origin* 41, 14–26. <https://doi.org/10.1016/j.gde.2016.07.003>
- Schuenemann, V.J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mitnik, A., Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J.D., White, W., Krause, J., Poinar, H.N., 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proceedings of the National Academy of Sciences* 108, E746–E752. <https://doi.org/10.1073/pnas.1105107108>
- Schuenemann, V.J., Peltzer, A., Welte, B., van Pelt, W.P., Molak, M., Wang, C.-C., Furtwängler, A., Urban, C., Reiter, E., Nieselt, K., Teßmann, B., Francken, M., Harvati, K., Haak, W., Schiffels, S., Krause, J., 2017. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications* 8, 15694. <https://doi.org/10.1038/ncomms15694>
- Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J.L., Kjellström, A., Wu, H., Stewart, G.R., Taylor, G.M., Bauer, P., Lee, O.Y.-C., Wu, H.H.T., Minnikin, D.E., Besra, G.S., Tucker, K., Roffey, S., Sow, S.O., Cole, S.T., Nieselt, K., Krause, J., 2013. Genome-Wide Comparison of Medieval and Modern *Mycobacterium leprae*. *Science* 341, 179–183. <https://doi.org/10.1126/science.1238286>
- Seymour, G.J., Ford, P.J., Cullinan, M.P., Leishman, S., Yamazaki, K., 2007. Relationship between periodontal infections and systemic disease. *Clinical Microbiology and Infection* 13, 3–10. <https://doi.org/10.1111/j.1469-0691.2007.01798.x>
- Shillington, K., 2012. *History of Africa*, Third. ed. Palgrave MacMillan, New York.
- Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., Rasilla, M. de la, Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M.V., Miller, R., Stewart, J.R., Derevianko, A.P., Jacobs, Z., Li, B., Roberts, R.G., Shunkov, M.V., Lumley, H. de, Perrenoud, C., Gušić, I., Kućan, Ž., Rudan, P., Aximu-Petri, A., Essel, E., Nagel, S., Nickel, B., Schmidt, A., Prüfer, K., Kelso, J., Burbano, H.A., Pääbo, S., Meyer, M., 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* eaam9695. <https://doi.org/10.1126/science.aam9695>
- Socransky, S. s., Haffajee, A. d., Cugini, M. a., Smith, C., Kent, R.L., 1998. Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology* 25, 134–144. <https://doi.org/10.1111/j.1600-051X.1998.tb02419.x>
- Socransky, S.S., Haffajee, A.D., 2002. Dental biofilms: difficult therapeutic targets. *Periodontology* 2000 28, 12–55.
- Søe, M.J., Nejsum, P., Seersholm, F.V., Fredensborg, B.L., Habraken, R., Haase, K., Hald, M.M., Simonsen, R., Højlund, F., Blanke, L., Merkyte, I., Willerslev, E., Kapel, C.M.O., 2018. Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. *PLOS ONE* 13, e0195481. <https://doi.org/10.1371/journal.pone.0195481>
- Sparacello, V.S., Roberts, C.A., Kerudin, A., Müller, R., 2017. A 6500-year-old Middle Neolithic child from Pollera Cave (Liguria, Italy) with probable multifocal osteoarticular tuberculosis. *International Journal of Paleopathology*. <https://doi.org/10.1016/j.ijpp.2017.01.004>
- Spigelman, M., Lemma, E., 1993. The use of the polymerase chain reaction (PCR) to detect *Mycobacterium tuberculosis* in ancient skeletons. *International Journal of Osteoarchaeology* 3, 137–143. <https://doi.org/10.1002/oa.1390030211>

- Spigelman, M., Matheson, C., Lev, G., Greenbaltt, C., Donoghue, H.D., 2002. Confirmation of the presence of Mycobacterium-tuberculosis complex-specific DNA in three archaeological specimens. *International Journal of Osteoarchaeology* 12, 393–401.
- Spyrou, M.A., Keller, M., Tikhbatova, R.I., Nelson, E.A., Valtueña, A.A., Walker, D., Alterauge, A., Carty, N., Fetz, H., Gourvenec, M., Hartle, R., Henderson, M., Heyking, K. von, Kacki, S., Knox, E.L., Later, C., Peters, J., Schreiber, J., Castex, D., Löscher, S., Harbeck, M., Herbig, A., Bos, K.I., Krause, J., 2018a. A phylogeography of the second plague pandemic revealed through the analysis of historical *Y. pestis* genomes. *bioRxiv* 481242. <https://doi.org/10.1101/481242>
- Spyrou, M.A., Tikhbatova, R.I., Wang, C.-C., Valtueña, A.A., Lankapalli, A.K., Kondrashin, V.V., Tsybin, V.A., Khokhlov, A., Kühnert, D., Herbig, A., Bos, K.I., Krause, J., 2018b. Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nature Communications* 9, 2234. <https://doi.org/10.1038/s41467-018-04550-9>
- Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *PNAS* 94, 9869–9874.
- Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A.J., 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *PNAS* 110, 228–233. <https://doi.org/10.1073/pnas.1207965110>
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., Fenner, L., Rutaiwa, L., Borrell, S., Luo, T., Gao, Q., Kato-Maeda, M., Ballif, M., Egger, M., Macedo, R., Mardassi, H., Moreno, M., Vilanova, G.T., Fyfe, J., Globan, M., Thomas, J., Jamieson, F., Guthrie, J.L., Asante-Poku, A., Yeboah-Manu, D., Wampande, E., Ssengooba, W., Joloba, M., Boom, W.H., Basu, I., Bower, J., Saraiva, M., Vasconcellos, S.E.G., Suffys, P., Koch, A., Wilkinson, R., Gail-Bekker, L., Malla, B., Ley, S.D., Beck, H.-P., de Jong, B.C., Toit, K., Sanchez-Padilla, E., Bonnet, M., Gil-Brusola, A., Frank, M., Penlap Beng, V.N., Eisenach, K., Alani, I., Ndung'u, P.W., Revathi, G., Gehre, F., Akter, S., Ntoumi, F., Stewart-Isherwood, L., Ntinginya, N.E., Rachow, A., Hoelscher, M., Cirillo, D.M., Skenders, G., Hoffner, S., Bakonyte, D., Stakenas, P., Diel, R., Crudu, V., Moldovan, O., Al-Hajoj, S., Otero, L., Barletta, F., Carter, E.J., Diero, L., Supply, P., Comas, I., Niemann, S., Gagneux, S., 2016. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat Genet* 48, 1535–1543. <https://doi.org/10.1038/ng.3704>
- Tellier, A., Lemaire, C., 2014. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *Molecular Ecology* 23, 2637–2652. <https://doi.org/10.1111/mec.12755>
- Tito, R.Y., Knights, D., Metcalf, J., Obregon-Tito, A.J., Cleeland, L., Najar, F., Roe, B., Reinhard, K., Sobolik, K., Belknap, S., Foster, M., Spicer, P., Knight, R., Jr, C.M.L., 2012. Insights from Characterizing Extinct Human Gut Microbiomes. *PLOS ONE* 7, e51146. <https://doi.org/10.1371/journal.pone.0051146>
- Tito, R.Y., Macmil, S., Wiley, G., Najar, F., Cleeland, L., Qu, C., Wang, P., Romagne, F., Leonard, S., Ruiz, A.J., Reinhard, K., Roe, B.A., Lewis, C.M., 2008. Phylotyping and Functional Analysis of Two Ancient Human Microbiomes. *PLoS ONE* 3, e3703. <https://doi.org/10.1371/journal.pone.0003703>
- Vågene, Å.J., Herbig, A., Campana, M.G., García, N.M.R., Warinner, C., Sabin, S., Spyrou, M.A., Valtueña, A.A., Huson, D., Tuross, N., Bos, K.I., Krause, J., 2018. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution* 520–528. <https://doi.org/10.1038/s41559-017-0446-6>
- Vandeventer, P.E., Lin, J.S., Zwang, T.J., Nadim, A., Johal, M.S., Niemz, A., 2012. Multiphasic DNA Adsorption to Silica Surfaces under Varying Buffer, pH, and Ionic

- Strength Conditions. *J Phys Chem B* 116, 5661–5670.
<https://doi.org/10.1021/jp3017776>
- Velsko, I.M., Frantz, L.A.F., Herbig, A., Larson, G., Warinner, C., 2018. Selection of Appropriate Metagenome Taxonomic Classifiers for Ancient Microbiome Research. *mSystems* 3, e00080-18. <https://doi.org/10.1128/mSystems.00080-18>
- Verma, D., Garg, P.K., Dubey, A.K., 2018. Insights into the human oral microbiome. *Arch Microbiol* 200, 525–540. <https://doi.org/10.1007/s00203-018-1505-3>
- Wagner, D.M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J.W., Enk, J., Birdsell, D.N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M., Golding, B., Riehm, J.M., Earn, D.J.D., DeWitte, S., Rouillard, J.-M., Grupe, G., Wiechmann, I., Bliska, J.B., Keim, P.S., Scholz, H.C., Holmes, E.C., Poinar, H., 2014. *Yersinia pestis* and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet Infectious Diseases* 14, 319–326. [https://doi.org/10.1016/S1473-3099\(13\)70323-2](https://doi.org/10.1016/S1473-3099(13)70323-2)
- Waldron, T., 2009. *Palaeopathology*. Cambridge University Press, Cambridge, UK.
- Warinner, C., Herbig, A., Mann, A., Yates, J.A.F., Weiß, C.L., Burbano, H.A., Orlando, L., Krause, J., 2017. A Robust Framework for Microbial Archaeology. *Annual Review of Genomics and Human Genetics* 18, null. <https://doi.org/10.1146/annurev-genom-091416-035526>
- Warinner, Christina, Rodrigues, J.F.M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H.U., Salazar-García, D.C., Eppler, E., Seiler, R., Hansen, L.H., Castruita, J.A.S., Barkow-Oesterreicher, S., Teoh, K.Y., Kelstrup, C.D., Olsen, J.V., Nanni, P., Kawai, T., Willerslev, E., von Mering, C., Lewis, C.M., Collins, M.J., Gilbert, M.T.P., Rühli, F., Cappellini, E., 2014. Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics* 46, 336–344. <https://doi.org/10.1038/ng.2906>
- Warinner, C., Speller, C., Collins, M.J., Lewis, C.M., 2015. Ancient human microbiomes. *Journal of Human Evolution* 79, 125–136. <https://doi.org/10.1016/j.jhevol.2014.10.016>
- Warinner, Christina, Speller, C., Collins, M.J., 2014. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, 20130376–20130376. <https://doi.org/10.1098/rstb.2013.0376>
- White, D.J., 1997. Dental calculus: recent insights into occurrence, formation, prevention, removal and oral health effects of supragingival and subgingival deposits. *European journal of oral sciences* 105, 508–522.
- White, D.J., 1991. Processes contributing to the formation of dental calculus. *Biofouling* 4, 209–218. <https://doi.org/10.1080/08927019109378211>
- WHO, 2018. WHO | Tuberculosis (TB) [WWW Document]. WHO. URL <http://www.who.int/tb/en/> (accessed 11.18.18).
- Wilbur, A.K., Bouwman, A.S., Stone, A.C., Roberts, C.A., Pfister, L.-A., Buikstra, J.E., Brown, T.A., 2009. Deficiencies and challenges in the study of ancient tuberculosis DNA. *Journal of Archaeological Science* 36, 1990–1997. <https://doi.org/10.1016/j.jas.2009.05.020>
- Wilson, I.G., 1997. Inhibition and Facilitation of Nucleic Acid Amplification 63, 11.
- Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., Soolingen, D. van, Rüsche-Gerdes, S., Locht, C., Brisse, S., Meyer, A., Supply, P., Niemann, S., 2008. Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. *PLOS Pathogens* 4, e1000160. <https://doi.org/10.1371/journal.ppat.1000160>
- Wood, J.R., Crown, A., Cole, T.L., Wilmschurst, J.M., 2016. Microscopic and ancient DNA profiling of Polynesian dog (kurī) coprolites from northern New Zealand. *Journal of Archaeological Science: Reports* 6, 496–505. <https://doi.org/10.1016/j.jasrep.2016.03.020>
- Wood, J.R., Rawlence, N.J., Rogers, G.M., Austin, J.J., Worthy, T.H., Cooper, A., 2008. Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). *Quaternary Science Reviews, Ice Age Refugia and Quaternary Extinctions: An Issue of Quaternary Evolutionary Palaeoecology* 27, 2593–2602. <https://doi.org/10.1016/j.quascirev.2008.09.019>

- Wood, J.R., Wilmshurst, J.M., Worthy, T.H., Cooper, A., 2012. First coprolite evidence for the diet of *Anomalopteryx didiformis*, an extinct forest ratite from New Zealand. *New Zealand Journal of Ecology* 36, 7.
- Xia, Y., Sun, J., Chen, D.-G., 2018. *Statistical analysis of microbiome data with R*. Springer Berlin Heidelberg, New York, NY.
- Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.-H., 2011. Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection. *Syst Biol* 60, 150–160. <https://doi.org/10.1093/sysbio/syq085>
- Yano, J.M., Yu, K., Donaldson, G.P., Shastri, G.G., Ann, P., Ma, L., Nagler, C.R., Ismagilov, R.F., Mazmanian, S.K., Hsiao, E.Y., 2015. Indigenous Bacteria from the Gut Microbiota Regulate Host Serotonin Biosynthesis. *Cell* 161, 264–276. <https://doi.org/10.1016/j.cell.2015.02.047>
- Yeh, H.-Y., Pluskowski, A., Kalējs, U., Mitchell, P.D., 2014. Intestinal parasites in a mid-14th century latrine from Riga, Latvia: fish tapeworm and the consumption of uncooked fish in the medieval eastern Baltic region. *Journal of Archaeological Science* 49, 83–89. <https://doi.org/10.1016/j.jas.2014.05.001>
- Yeh, H.-Y., Prag, K., Clamer, C., Humbert, J.-B., Mitchell, P.D., 2015. Human intestinal parasites from a Mamluk Period cesspool in the Christian quarter of Jerusalem: Potential indicators of long distance travel in the 15th century AD. *International Journal of Paleopathology* 9, 69–75. <https://doi.org/10.1016/j.ijpp.2015.02.003>
- Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M.J., Langridge, G., Fotakis, A.K., Nair, S., Stenøien, H.K., Hamre, S.S., Casjens, S., Christophersen, A., Quince, C., Thomson, N.R., Weill, F.-X., Ho, S.Y.W., Gilbert, M.T.P., Achtman, M., 2018. Pan-genome Analysis of Ancient and Modern *Salmonella enterica* Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Current Biology* 28, 2420–2428.e10. <https://doi.org/10.1016/j.cub.2018.05.058>
- Ziesemer, K.A., Mann, A.E., Sankaranarayanan, K., Schroeder, H., Ozga, A.T., Brandt, B.W., Zaura, E., Waters-Rist, A., Hoogland, M., Salazar-García, D.C., Aldenderfer, M., Speller, C., Hendy, J., Weston, D.A., MacDonald, S.J., Thomas, G.H., Collins, M.J., Lewis, C.M., Hofman, C., Warinner, C., 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports* 5, 16498. <https://doi.org/10.1038/srep16498>
- Zink, A. R., Grabner, W., Reischl, U., Wolf, H., Nerlich, A.G., 2003. Molecular study on human tuberculosis in three geographically distinct and time delineated populations from ancient Egypt. *Epidemiology & Infection* 130, 239–249. <https://doi.org/10.1017/S0950268802008257>
- Zink, A.R., Haas, C.J., Reischl, U., Szeimies, U., Nerlich, A.G., 2001. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *Journal of Medical Microbiology* 50, 355–366. <https://doi.org/10.1099/0022-1317-50-4-355>
- Zink, Albert R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H., Nerlich, A.G., 2003. Characterization of *Mycobacterium tuberculosis* Complex DNAs from Egyptian Mummies by Spoligotyping. *Journal of Clinical Microbiology* 41, 359–367. <https://doi.org/10.1128/JCM.41.1.359-367.2003>

9. Summary

Ancient microbial DNA is a powerful source of information about past human health, and our ability to effectively analyze and utilize this information is still nascent. Next generation sequencing technologies have allowed us to extract complete bacterial genomes from the large microbial communities that inhabit ancient human remains (or associated archaeological materials) as well as access the communities themselves. This dissertation illustrates the value in both approaches through the metagenomic analysis of diverse archaeological materials: a calcified lung nodule from mummified remains, dentin, dental calculus, and bulk latrine sediments. Over the three projects presented here, I utilize a 17th century genome from the *Mycobacterium tuberculosis* complex (MTBC) to better calibrate the emergence of the complex, present preservation dynamics of host and oral microbiome DNA in a global dataset of paired dentin and dental calculus, and explore microbial DNA preservation in bulk latrine samples in terms of both eukaryotic parasites and gut-associated microbes.

Manuscript A presents the reconstruction of a high-coverage *Mycobacterium tuberculosis* genome. The high quality of the genome and the precision of the year of death for the deceased person allowed us to use the genome as a confident calibration point in Bayesian phylogenetic analysis to date the time to the most recent common ancestor (tMRCA) for the full MTBC and Lineage 4 of the MTBC. This project demonstrates that calcified lung nodules are an excellent source for human and MTBC DNA, and supports the narrative that tuberculosis as we know it today emerged during the spread of agriculture and animal domestication at the earliest.

Manuscript B presents a comprehensive exploration of DNA preservation in dental calculus (calcified dental plaque) when compared to dentin samples from the same tooth, systematically confirming that dental calculus provides an excellent preservation environment for DNA across different continents, time periods, and individuals. All calculus samples provided strong genetic signals from oral microbes and low consistent levels of host DNA, seemingly robust against environmental contamination. Intriguingly, dentin was also found to stochastically harbor low levels of representation from oral microbes, indicating it may be used as a proxy for individual oral taxa in lieu of dental calculus in some situations. This project sets a standard for universal expectations of DNA preservation in dental calculus, which had not been available prior to its publication.

Manuscript C presents a case study in interdisciplinary paleomicrobiology, in which bulk latrine sediments from two medieval sites which had been previously analyzed by microscopy are genetically analyzed. The metagenomic profiles of the two sites are investigated for signals of eukaryotic parasites and community gut microbiome preservation.

This project illustrates how metagenomic analysis can complement microscopy in investigating community health in the past by offering in theory a more specific and fine-grained view of not only what taxa were present in the latrine but their lineage and genomic sequences, as well as the invisible world of gut-associated microbial taxa. However, this project also highlights many limitations to analyses of this kind, most of which stem from a paucity of modern genomic sequences for eukaryotic parasites and unculturable gut microbes.

Overall, this dissertation explores the value and limitations of ancient microbial genetic data in exploring human health and disease in the past.

10. Zusammenfassung

Alte mikrobielle DNA ist eine ausgezeichnete Quelle um Informationen über die Gesundheit der Menschen in der Vergangenheit zu gewinnen. Die Methoden zur effektiven Analyse sind noch jung und müssen weiter präzisiert werden. Next-Generation-Sequenzierungstechnologien machten es möglich, vollständige Bakteriengenome aus großen mikrobiellen Gemeinschaften menschlicher Überreste zu isolieren sowie die Gemeinschaften in ihrer Gesamtheit zu untersuchen. Die vorgelegte Dissertation veranschaulicht den Wert dieser beiden Ansätze durch die metagenomische Analyse verschiedener archäologischer Materialien: ein verkalkter Lungenknoten aus mumifizierten menschlichen Überresten, Dentin, Zahnstein und Latrinsedimenten. Im Verlauf der drei hier vorgestellten Projekte benutze ich das Genom eines *Mycobacterium tuberculosis* complex (MTBC) aus dem 17. Jahrhundert um den Zeitpunkt der Entstehung des Komplexes besser zu kalibrieren, untersuche anhand eines globalen Datensatzes die Dynamik in der DNA Erhaltung von Wirt und oralem Mikrobiom in Zahnstein und Zahnbein und untersuche die Konservierung mikrobieller DNA in zwei Latrinsedimentenproben für sowohl eukaryotische Parasiten als auch im Darm angesiedelte Mikroben.

Manuskript A behandelt ein hochqualitatives *Mycobacterium tuberculosis*-Genom. Die hohe Qualität des Genoms und das genau bekannte Todesjahr des Verstorbenen ermöglichen es uns, das Genom als sicheren Kalibrierungspunkt in der Bayesischen phylogenetischen Analyse zu verwenden und so den letzten gemeinsamen Vorfahren (tMRCA) für den gesamten MTBC und der Linie 4 des MTBC zu bestimmen. Diese Arbeit zeigt, dass der verkalkte Lungenknoten eine ausgezeichnete Quelle für MTBC-DNA darstellt und unterstützt die These der Verbindung zwischen der Entstehung der Tuberkulose und der Ausbreitung der Landwirtschaft und Tierzucht.

Manuskript B bietet eine umfassende Untersuchung der DNA-Erhaltung in Zahnstein im Vergleich zum Zahnbein desselben Zahns und bestätigt systematisch Zahnstein als eine hervorragende Konservierungsumgebung für DNA über verschiedene Kontinente, Zeiträume und Individuen hinweg. Alle Zahnsteinproben weisen einen hohen Gehalt an genetischem Material der oralen Mikroben und geringen Mengen an Wirts DNA auf. Die Zahnsteinproben enthielten außerdem weniger Verunreinigungen aus der Umgebung. Interessanterweise enthielt auch das Zahnbein kleine Mengen von oralen Mikroben. Dies zeigt, dass Dentin in einigen Situationen als Proxy für einzelne orale Taxa verwendet werden kann. Diese Arbeit erhebt damit standardisiert die Erwartungen für die Erhaltung von DNA in Zahnstein.

In Manuskript C wird eine Fallstudie in der interdisziplinären Paläomikrobiologie vorgestellt. Latrinsedimentproben aus zwei mittelalterlichen Stätten, die zuvor bereits mikroskopisch analysiert wurden, werden genetisch analysiert. Die metagenomischen Profile

beider Proben wurden auf eukaryotischen Parasiten und Erhaltung des Darm- Mikrobioms hin untersucht. Das Projekt zeigt, wie die Metagenomanalyse die Mikroskopie bei der Untersuchung der Gesundheit vergangener Bevölkerungen ergänzen kann, indem sie einen genaueren Einblick in das Vorhandensein von Taxa, ihre Abstammungslinie und ihre Genomsequenzen bietet. Genetische Analysen können auch die unsichtbare Welt der zum Darmmikrobiom gehörenden Taxa aufdecken. Allerdings zeigt dieses Projekt auch viele Einschränkungen bei Analysen dieser Art auf. Die meisten Einschränkungen ergeben sich aus dem Mangel an modernen Genomsequenzen für eukaryotische Parasiten und nicht kultivierbarer Darmmikroben.

Zusammenfassend untersucht diese Arbeit die Möglichkeiten und Grenzen alter mikrobieller genetischer Daten bei der Erforschung der menschlichen Gesundheit und Krankheiten der Vergangenheit.

11. Eigenständigkeitserklärung

Entsprechend §5 Abs. 4 der Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena, erkläre ich, dass mir die geltende Promotionsordnung der Fakultät bekannt ist. Ich bezeuge, dass ich die vorliegende Dissertation selbst angefertigt habe und keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen sowie Quellen in meiner vorliegenden Arbeit angegeben habe. Zudem habe ich alle Personen, die mir bei der Auswahl und Auswertung sowie bei der Erstellung der Manuskripte unterstützt haben, in der Auflistung der Manuskripte und den entsprechenden Danksagungen namentlich erwähnt. Zudem versichere ich, dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen haben und auch Dritten von mir keine unmittelbaren sowie mittelbaren geldwerte Leistungen für Arbeiten, die im Zusammenhang mit dieser Dissertation stehen, erhalten haben. Die vorliegende Promotion wurde zuvor weder für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, also auch einer anderen Hochschule als Dissertation vorgelegt.

Jena, den 24.06.2019

Susanna Sabin

12. Curriculum Vitae

CURRICULUM VITAE

Susanna Jacey Sabin

sabin@shh.mpg.de | susanna.sabin@gmail.com

+49 (0) 3641 686-653

Max Planck Institute for the Science of Human History
Department of Archaeogenetics
Kahlaische Straße 10
07745 Jena, DE

Education

- 2016 – present **Max Planck Institute for the Science of Human History, Molecular Palaeopathology, Ph.D.** | To be awarded by Friedrich-Schiller-Universität Jena, Faculty of Biology and Pharmacy | Expected October 2019 | Jena, Germany
Advisors: Prof. Dr. Johannes Krause, Dr. Kirsten Bos
Dissertation Topic: Insights into microbial evolution and ecology from genetic analysis of diverse archaeological materials
- 2014 – 2016 **Universität Tübingen, Archaeological Sciences, M.Sc.** | Tübingen, Germany
Advisors: Prof. Dr. Johannes Krause, Dr. Kirsten Bos
Thesis: Ancient DNA analysis of dental remains from Kilteasheen: a case study in metagenomics and an exploration of dental calculus
Description: Screening of paired archaeological dentine and dental calculus samples from the medieval site of Kilteasheen in Ireland for human and microbial DNA preservation revealed key differences in preservation between the materials over 49 dentine and 36 calculus samples.
Note: 1,1 "Sehr gut"
- 2010 – 2014 **California State University Los Angeles, Anthropology, B.A.** | *Magna Cum Laude* | Los Angeles, CA, USA
Advisors: Dr. Elizabeth Miller, Dr. Matthias Selke
Thesis: Tracing the origin of tuberculosis: an integrated paleopathological perspective
Description: This thesis addresses existing evidence and proposes a unified hypothesis for the origin of tuberculosis in human populations, with an emphasis on the need for future interdisciplinary cooperation.
GPA: 3.76
- 2012 – 2013 **University of St Andrews, Junior Year Abroad** | St. Andrews, Scotland, UK

Publications

- 2019 Sabin, S., Herbig, A., Vågane, Å.J., Ahlström, T., Bozovic, G., Arcini, C., Kühnert, D., Bos, K.I. (submitted). A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex. Preprint available on bioRxiv: <https://doi.org/10.1101/588277>
- 2019 Key, F.M., Posth, C., Esquivel-Gomez, L.R., Huebler, R., Spyrou, M., Neumann, G., Furtwaengler, A., **Sabin, S.**, Burri, M., Wissgott, A., et al. (in review). 6,500 year old *Salmonella enterica* genomes link human-host adaptation to animal domestication.
- 2019 Bos, K.I., Kühnert, D., Herbig, A., Esquivel-Gomez, L.R., Andrades Valtueña, Barquera, R., Giffin, K., Kumar-Lankapalli, A., Nelson, E.A., **Sabin, S.**, Spyrou, M.A., Krause, J. (in press). Paleomicrobiology: Diagnosis and Evolution of Ancient Pathogens. *Annu. Rev. Microbiol.*
- 2018 Mann, A.*, **Sabin, S.***, Ziesemer, K., Vågane, Å.J., Schroeder, H., Ozga, A.T., Sankaranarayanan, K., Hofman, C.A., Fellows Yates, J.A., Salazar-García, D.C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G.R., Stone, A.C., Lewis, C.M., Krause, J., Hofman, C., Bos, K., Warinner, C. 2018. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific Reports* 8, 9822. doi: 10.1038/s41598-018-28091-9.
- 2018 Vågane, Å.J., Herbig, A., Campana, M.G., García, N.M.R., Warinner, C., **Sabin, S.**, Spyrou, M.A., Valtueña, A.A., Huson, D., Tuross, N., Bos, K.I., Krause, J., 2018. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecol. Evo.* 1. <https://doi.org/10.1038/s41559-017-0446-6>

* Equal contributions

Conference Posters/Presentations

- 2018 **Sabin, S.** "The Bishop Returns, and other stories..." Talk presented for the institute-wide Work In Progress Seminar at the MPI for the Science of Human History in Jena, Germany on 11 October 2018.
- 2018 **Sabin, S.**, Herbig, A., Kühnert, D., Vågane, Å.J., Ahlström, T., Bozovic, G., Nelson, E.A., Arcini, C., Bos, K.I. "Analysis of a 17th century *Mycobacterium tuberculosis* genome from Lund, Sweden extracted from a lung nodule." Poster presented at the International Symposium for Biomolecular Archaeology in Jena, Germany on 19 September 2018.
- 2018 **Sabin, S.**, Herbig, A., Vågane, Å.J., Ahlström, T., Bozovic, G., Nelson, E.A., Arcini, C., Bos, K.I. "A lung nodule as a rich source of host and pathogen DNA: Analysis of a 17th century *Mycobacterium tuberculosis* genome from Lund, Sweden." Talk presented at the European Palaeopathology Association meeting in Zagreb, Croatia on 31 August 2018.
- 2018 **Sabin, S.**, Herbig, A., Vågane, Å.J., Ahlström, T., Bozovic, G., Nelson, E.A., Arcini, C., Bos, K. "A high-quality 17th century *Mycobacterium tuberculosis* genome adds resolution to mycobacterial dating and phylogeography." Talk presented during the symposium "Tracing back bacterial pathogen evolution

from ancient and modern genomics” at the Society for Molecular Biology and Evolution meeting in Yokohama, Japan on 08 July 2018.

- 2018** **Sabin, S.**, Herbig, A., Vågene, Å.J., Ahlström, T., Bozovic, G., Nelson, E.A., Arcini, C., Bos, K. “Application of a 17th Century *Mycobacterium tuberculosis* genome to mycobacterial dating and phylogeography.” Poster and rapid fire talk presented at the American Society for Microbiology Microbe meeting in Atlanta, GA, USA on 10 June 2018. Recipient of Outstanding Abstract Award.
- 2016** **Sabin, S.**, Vågene, Å., Fellows Yates, J., Read, C., Herbig, A., Krause, J., Bos, K. “A Comparative Dataset of Dentine and Dental Calculus from Medieval Ireland.” Poster presented at the 7th International Symposium on Biomolecular Archaeology (ISBA7) in Oxford, UK on 14-16 September 2016.
- 2014** **Sabin, S.** “Tracing the Origin of Tuberculosis: An integrated paleopathological perspective.” Podium presentation presented at the California State University Honors Consortium Conference in Fullerton, CA, USA on 3 May 2014.
- 2014** **Sabin, S.** “Tracing the Origin of Tuberculosis: An integrated paleopathological perspective.” Podium presentation presented at the Southwestern Anthropological Association Conference (SWAA) in Anaheim, CA, USA on 26 April 2014.

Honors/Awards

- 2018** American Society for Microbiology Outstanding Abstract Award (200 USD)
- 2017** Wenner-Gren Dissertation Fieldwork Grant (19,385 USD)
- 2014** CSULA Honors College Scholarship (2,000 USD)
- 2014** The Honor Society of Phi Kappa Phi
- 2012 – 2014** Lambda Alpha National Anthropological Honor Society
- 2012 – 2013** CSULA Dean’s List in recognition of superior scholastic achievement
- 2012 – 2013** Special Recognition while in the Honors College
- 2011** Special Recognition while in the General Education Honors Program

Field Experience

- 2015** **Student Excavator** | Palaeolithic site complex at Schöningen, Germany
Supervisor: Dr. Jordi Serengali, Universität Tübingen
Note: 1,0
- 2012** **Student Excavator** | Koobi Fora Field School | Turkana Basin, Kenya
Supervisors: Dr. John W.K. Harris, Rutgers University | Dr. David Braun, University of Cape Town | Dr. Purity Kiura, National Museums of Kenya
Note: A

Service/Engagement

- 2019** Reviewer for the *American Journal of Physical Anthropology*
- 2018-2019** PhD representative on the managing board, International Max Planck Research School for the Science of Human History
- 2018-2019** Elected external PhD representative for the Max Planck Institute for the Science of Human History; Jena, Germany
- 2018** Department of Archaeogenetics journal club coordinator

- 2018 Co-author of Dental Calculus Large Specimen Sampling Protocol; Resource for archaeologists (<http://christinawarinner.com/resources/archaeologists/>)
- 2017 Volunteer for Lange Nacht der Wissenschaften Jena (Long Night of the Sciences, Jena); Max Planck Institute for the Science of Human History; Jena, Germany
- 2014 Elder council for the Early Entrance Program provisional summer program at CSULA; Los Angeles, CA
- 2013 Head mentor for the Early Entrance Program provisional summer program at CSULA; Los Angeles, CA
- 2011 – 2012 Volunteer in Pathology Department; Glendale Memorial Hospital; Glendale, CA
- 2011 Mentor for the Early Entrance Program provisional summer program; Los Angeles, CA

Work Experience

- 2016 - present **Doctoral candidate** | Molecular Palaeopathology group, Department of Archaeogenetics, Max Planck Institute for the Science of Human History
Supervisor: Dr. Kirsten Bos
- 2015 - 2016 **HiWi (Research Assistant)** | Department of Archaeogenetics, Max Planck Institute for the Science of Human History
Supervisors: Prof. Dr. Johannes Krause, Dr. Kirsten Bos, Åshild Vågane
- 2014 – 2015 **HiWi (Research Assistant)** | Paleogenetics group, Institute for Archaeological Sciences, Universität Tübingen
Supervisors: Prof. Dr. Johannes Krause, Dr. Kirsten Bos, Åshild Vågane
- 2013 – 2014 **Student Technology Assistant** | Center for Effective Teaching and Learning | California State University, Los Angeles
Supervisors: Catharine Haras, Dr. Beverly Bondad-Brown
- 2013 **Discussion Guide Author** | *GO PUBLIC: A Day in the Life of an American School District* (Documentary Film)
Supervisors: Dawn O’Keeffe, James W. O’Keeffe

Training & Supervision

- 2017 Technical training of PhD students (x2) in
 - Ancient DNA extraction (“Dabney” method)
 - Illumina library preparation

13. Acknowledgments

I want to first thank Prof. Johannes Krause and Dr. Kirsten I. Bos for allowing me to join the Molecular Paleopathology group at the Max Planck Institute for the Science of Human History, and giving me the opportunity to build my knowledge and skill set in archaeogenetics with compelling projects.

I would like to specifically thank Dr. Denise Kühnert, Dr. Alexander Herbig, Felix Key, Luis Esquivel Gomez, Aida Andrades Valtuena, Åshild J. Vågane, and James Fellows Yates for training and assistance with the computational work performed for this dissertation and more, and thank in general the PIs, postdocs, and PhD candidates at the MPI-SHH for continuous constructive feedback and support.

I also extend my gratitude to the members of our IT department at the MPI-SHH, without whom barely any of this work could have been performed, and the incredible administrative team in the Department of Archaeogenetics and the MPI-SHH for their support.

I was supported in the production of this document by suggestions from Prof. Johannes Krause and Dr. Kirsten I. Bos. Kathrin Nägele provided invaluable assistance in translating my summary to German, and my family helped me with proof-reading. Thank you also to Kathy Anderson, my Dad, and everyone else in my life who taught me how to write.

On a more personal note:

Thank you first to my fellow PhD students at the MPI-SHH. Thank you second to all my singing/guitar-playing/piano-playing people in the Department of Archaeogenetics who joined me in perhaps a-little-too-enthusiastically covering pop songs at various office parties, retreats, and Karaoke nights.

Thank you to the Kuppakuolema crew (you know who you are), Pocari the hamster, and the producers of Pocari Sweat (elixir of life). Our Japanese adventures will hold a special place in my heart forever, and I will surely annoy my grandchildren with tales of them. I love you guys, and even if I fling myself back to the deserts of North America, you can't get rid of me that easy. Except, pretend I expressed that in a less-creepy way.

Thank you to my parents for once again supporting me even when my endeavors took me across the planet. I'm looking forward to movie days, weekend breakfasts, and trips to Starbucks being only a slightly uncomfortable drive and a good audiobook away.

To Elizabeth Nelson – my roomie, my officemate, my commiserative person, my distinguished colleague and friend, who convinced me to go to the emergency room that one time – I would just like to say:

[Jack Lemmon yelling aggressively at Walter Matthau] “You’re a wonderful guy, Oscar! You’ve done everything for me, if it weren’t for you I don’t know what would’ve happened to me. You took me in here, you gave me a place to live, you gave me something to live for. I’m never gonna forget you for that Oscar. You’re TOPS with me.” (Simon, N. 1968. “The Odd Couple.” Saks, G. [Dir.], Koch, H.W. [Prod.]

And finally, to Cliff – thank you for choosing me back and being the home I come back to.