UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

COMPARISON OF FIVE DIFFERENT DNA EXTRACTION METHODS FOR MICROBIOME RECONSTRUCTION FROM COPROLITES

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF ARTS

By

RICHARD WILLIAM HAGAN
Norman, Oklahoma
2017

COMPARISON OF FIVE DIFFERENT DNA EXTRACTION METHODS FOR MICROBIOME RECONSTRUCTION FROM COPROLITES

A THESIS APPROVED FOR THE
DEPARTMENT OF ANTHROPOLOGY

BY

_____
Dr. Christina Warinner, Chair

_____
Dr. Cecil M. Lewis, Jr.

_____
Dr. Krithivasan Sankaranarayanan

Acknowledgements

I am eternally grateful for the support and patience of my committee members, whose continued guidance and encouragement have been a sustaining force for my scientific goals. I thank Dr. Cecil M. Lewis, Jr. for the many lengthy discussions that have refined my understanding of evolution and biology in general. I thank Dr. Krithivasan "Krithi" Sankaranarayanan for his insistence that I understand the bioinformatic analyses I use, and for letting me learn from my mistakes. And of course I am exceptionally grateful to my Chair, Dr. Christina "Tina" Warinner, whose wide expertise and reasoned guidance continue to be instrumental to my development as a scientist. I thank the full cohort of faculty and staff of the Department of Anthropology who have supported, challenged, and encouraged me in my early career. I am especially grateful to Drs. Courtney Hofman and Stephanie Schnorr for their mentorship and critical feedback. Last but not least, I must thank my lab mates and some of the best friends and colleagues anyone could ever ask for. Rita Austin, Dave Jacobson, Justin Lund, Allie Mann, and Nisha Patel have been incredible sources of inspiration. Thank you for the countless favors, long nights together in the lab, and all of the laughter and tears along the way. This thesis is as much all of yours as it is mine.

Table of Contents

Acknowledgements iv

List of Tables vii

List of Figures viii

Abstract ix

Chapter 1: Background 1

Microbiomes 1

Technological changes and microbiome science 2

Current trends in microbiome science 4

Microbial ecology and the human microbiome 6

The human gut microbiome 9

Ancient DNA 11

Microbial Archaeology 16

Ancient microbiomes 17

Chapter 2: Coprolites 20

Coprolite studies to date 20

Lysis and DNA purification strategies with coprolites 21

The unknowns of aDNA from coprolites 22

The present study 23

Chapter 3: Materials and methods 24

Methods 24

Extraction methods 25

Extraction method 1: HMP protocol 25

Extraction method 2: HMP protocol with modified MinElute protocol 26
Extraction method 3: Phenol-chloroform with modified MinElute protocol 27
Extraction method 4: Modified MinElute column 28
Extraction method 5: Double modified MinElute column 28
DNA extract quantification 29
DNA library construction 29
Shotgun data analysis and quality filtering 31
Metataxonomic characterization 32
Bayesian source tracking of microbial communities 33
Ancient DNA validation, insert length analysis, and statistical testing 34
Results 34
DNA extraction yields 34
Reads 35
Metataxonomic characterization 35
Bayesian source tracking 36
Ancient DNA analysis 37
Fragment length analysis 37
Examining relationships between DNA damage, insert length, and experimental variables 38
Chapter 4: Discussion 53
References 57

List of Tables

Table 1. Normalized DNA yields 40

Table 2. Summary of reads 41

Table 3. Summary of reference genomes. 42

List of Figures

Figure 1. Normalized DNA yield by replicate 43

Figure 2. Weighted UniFraC β-diversity. 44

Figure 3. Microbial Community Source Tracking 45

Figure 4. Weighted UniFrac on SourceTracker data 46

Figure 5. Distribution of DNA damage by method and sample 47

Figure 6. Distribution of insert lengths by sample and method 48

Figure 7. Comparison of insert lengths by column type 49

Figure 8. Linear modeling of experimental variables 50

Figure 9. Variance of DNA damage by reference taxon 51

Figure 10. Variance of insert length by taxon 52

Abstract

The human microbiome has become an intense area of research over the past fifteen years due to its importance for understanding human health and disease. The revolution in Next-Generation Sequencing (NGS) techniques has made microbiome science accessible to many types of researchers, including biomolecular archaeologists, amongst whom interest in the ancestral state of the human microbiome is increasing. Two archaeological materials — dental calculus and coprolites — have been shown to harbor ancient microbial DNA that can be extracted, purified, and sequenced to reconstruct the oral and gut microbiomes of past peoples. While systematic experiments have improved upon techniques for the extraction of ancient DNA from dental calculus, comparable work has not been performed with coprolites. The goal of this study was to compare ancient DNA extraction and purification procedures using human coprolites that have previously yielded ancient DNA. Five DNA extraction methods were applied to the three coprolites, including a standard protocol utilizing a commercial fecal DNA isolation kit and four protocols progressively optimized for the recovery of ancient DNA. Each coprolite was subsampled to allow for the testing of all five extraction methods in duplicate, for a total of 30 extractions. The concentration of all DNA extracts was measured, and extracts were then converted into Next-Generation shotgun sequencing libraries. The libraries were pooled and sequenced, and a series of metataxonomic and statistical analyses were performed on the resultant data to test for differential impacts between extraction and purification chemistries. Methods designed specifically for ancient DNA recovered significantly more DNA than the tested

commercial kit; however, no significant shift in the microbial community structure of the samples was observed between extraction and purification strategies.

Chapter 1: Background

Microbiomes

Over the past fifteen years, the study of *microbiomes* has become an increasingly popular area of research. A commonly-cited fact is that Nobel laureate Joshua Lederberg coined and defined the term microbiome around 2001, in the article *'Ome Sweet 'Omics—A Geneaological Treasury of Words* (Lederberg and McCray 2001). The definition that Lederberg offers in that article has been interpreted as an extension of other –ome terms, such as genome or proteome (Huss 2014), that defines the collective genomes of all microorganisms living within a given niche. While this definition by Lederberg would eventually hold true, Lederberg's original definition is clearly ecological in nature, as he plainly refers to the microbiome as the "ecological community" of these microorganisms. It was not until 2004, in an article published in *New Perspectives Quarterly*, that Lederberg referred to the microbiome as the "cohabitation of genomes within the human body" (Lederberg 2004), supporting the term's relationship with other –omics.

Perhaps more interesting than the evolution of the meaning assigned to the word by its purported minter, however, is the fact that the term microbiome existed long before Lederberg began using it. As early as 1949, the term microbiome — albeit absent a definition — was used in the French journal *Revue Odontologique*, referring to how changes in pH can impact the evolution of the intestinal microbiome (Cambiés 1949). In the early 1950s, John L. Mohr of the Department of Zoology at the University of Southern California used the term in an article describing how changes in an aquatic microbiome could prove useful as indicators of pollution from industry (Mohr 1952).

Subsequent usage of the term microbiome through the remaining decades of the 20th century continued to focus mostly on its utility for describing small (micro) environments (biomes). The earliest explicit definition of the term microbiome in relation to microorganisms appears in the edited volume *Fungi in Biological Control Systems*, wherein a group of authors define it as "a characteristic microbial community occupying a reasonably well defined habitat…" (Whipps et al. 1988). This early definition, being ecological in nature, is much closer to Lederberg's first usage of the term.

Today, the term microbiome is commonly used to refer to the microorganisms that inhabit a given ecological niche, their collective genomes, the environmental conditions, or some combination thereof (Marchesi and Ravel 2015). In other words, the term microbiome is derived from the concept of a biome, and not the homonymous phoneme that is formed by appending –ome to the study of microbes, as is inferred from Lederberg's 2001 article.

Technological changes and microbiome science

The enduring legacy of Lederberg being credited with coining the term microbiome may be attributable to his timing at the dawn of the age of genomics. While authors prior to Lederberg used the term, the large-scale genomic study of the full cohort of microorganisms within a given niche was hampered by the technologies available to earlier researchers. It was not until 1977 that the Sanger method for sequencing DNA was published (Sanger et al. 1977), and although this new method for elucidating the sequence of DNA molecules revolutionized molecular biology, data generation using this technique was relatively slow. And despite pioneering research in

the field of collecting and annotating known protein and nucleic acid sequences being underway, particularly by Margaret Dayhoff (Eck and Dayhoff 1966), computational and throughput limitations presented a significant challenge for querying these resources.

Recognizing the need for a way to reliably and accurately distinguish organisms at the molecular level, Carl Woese developed the method of classifying microorganisms by the 16S subunit of their ribosomal RNA (rRNA) (Woese and Fox, 1977). Woese had previously shown that the primary structure of the 16S rRNA was a highly-conserved element due to its essential role in protein synthesis (Woese et al. 1975), and this information coupled with the emergence of then-rapid DNA sequencing technologies led to increased interest in this locus for resolving evolutionary relationships, resulting in the first complete 16S rRNA gene sequence of *E. coli* in 1978 (Brosius et al. 1978). While these advances were substantial, sequencing and classification of a gene of interest relied on having many copies of the molecule, typically obtained via the laborious practices of growing the bacteria of interest in pure culture, fragmenting their DNA with endonucleases (Danna and Nathans, 1971), cloning those fragments into plasmids and culturing again (Cohen et al. 1973), and eventually sequencing the amplified products. Performing this for every member of a complex microbiome was simply out of the question.

Two discoveries made roughly a decade after the publication of the 16S rRNA gene sequence of *E. coli* would dramatically streamline the identification of microorganisms by this locus. The first came in 1986, with the publication of the improved method of DNA amplification via the polymerase chain reaction (PCR)

(Mullis et al. 1986). The second was the publication of the so-called universal primers for the 16S rRNA gene (Weisburg et al. 1991). The combination of these approaches permitted for the first time the capability to single out and amplify most of the 16S rRNA genes from a microbiome; however, sequencing via capillary-based Sanger methods still presented a significant bottleneck to the downstream identification of these organisms.

As the 20th century was coming to a close, the final piece of the puzzle that would permit characterization of the entire microbial community within a sample was coming online. The invention of high-throughput DNA sequencing — so-called next-generation sequencing (NGS) — gave researchers the ability to rapidly generate orders of magnitude more sequences than previous methods. The concurrent development of mature sequence databases such as EMBL and GenBank, as well as fast and accurate sequence query tools such as BLAST (Altschul et al. 1990), provided the computational resources needed for comparing them. It was only then, at the confluence of these technologies, that Joshua Lederberg's definition of the term microbiome could find permanence amongst researchers who now had the tools at their disposal for studying these *small environments* and the organisms who constituted them.

Current trends in microbiome science

Since the publication of Lederberg's original article, the field of microbiome science has experienced tremendous growth. Most studies conducted thus far have focused on the portion of the microbiome that is constituted by all the organisms within an environment — the microbiota — but often refer to this portion simply by its parent

term microbiome. This confusion prompted the previously-referenced article from Marchesi and Ravel (2015), wherein the authors proposed a vocabulary for the field to decrease misunderstanding between researchers within it. For the present article, studies of the microbiome are assumed to be *metataxonomic* studies, or those which seek to characterize the members of a microbiome, most often by the characterization of the 16S rRNA genes present, but also through *metagenomics*, or characterization via all the genes and genomes present. Both approaches to metataxonomics have revealed a world of previously unrecognized functional dynamics at work between microbiota and the environments in which they are found.

One of the most compelling areas of microbiome research has been into the environmental niches that constitute the human body. Early recognition that this would be a fruitful topic for inquiry resulted in one of the largest collections of metataxonomic projects to date, the Human Microbiome Project (HMP) (Peterson et al. 2009), which sought to characterize the diversity of bacterial communities within and on the human body. Microbes, particularly bacteria, inhabit nearly every part of the human body, and differ between body sites in accordance with the fluctuating conditions present such as moisture, pH, and the oxygenic environment (Grice et al. 2009, Costello et al. 2009, Huttenhower et al. 2012). This difference in the composition of microbiomes across the human body has a strong analogical relationship with concepts from ecology, and indeed much of what is considered microbiome science today has its roots in the rapid changes that have taken place in the field of microbial ecology (Prosser et al. 2007).

Microbial ecology and the human microbiome

Amongst the ecological concepts most useful to the microbiome scientist are the measures α- and β-diversity, or the diversity within a community and the diversity between communities of microbes respectively (Costello et al. 2012). These measures allow for the quantitative and comparative studies of the microbiomes within and between groups of host organisms. Classical measures of α- and β-diversity in ecology rely on the ability to easily distinguish and count species within an environment. Due to the difficulty in defining microbes using most species concepts (Achtman et al. 2008, Philippot et al. 2010), they are typically binned into what are known as *operational taxonomic units* (OTUs) (Blaxter et al. 2005) before attempting to measure diversity. For the metataxonomic study of microbes, OTUs are typically created by clustering 16S rRNA gene sequences together at a defined similarity threshold, commonly 97%. Diversity and abundance can then be calculated on a wide range of formulae using these OTUs.

Alternatively, phylogenetic formulae can also be employed to calculate diversity. These formulae typically forgo the simple abundance or presence-absence data from classic ecology and instead exploit the phylogenetic information inherit in DNA sequencing data. For example, a commonly used measure for calculating the α-diversity of a microbiome is Faith's phylogenetic distance (Faith's PD) (Faith 1992), as it is robust to the effects of different taxonomic units or features. However, Faith's PD in its original form does not account for taxa (i.e. OTU) abundance, and thus other measures are recommended (McCoy and Matsen 2013). One such measure that has been developed for use in comparing communities between samples is UniFrac

(Lozupone and Knight 2005), which can provide both unweighted and weighted β-diversity estimates between samples using phylogenetic inference.

At the heart of these measures of diversity within and between samples lies the desire to be able to make meaningful inferences about observed differences in the community structure of microbiomes from different niches such as body sites. For example, it has been known since the days of Leeuwenhoek that the human oral cavity is inhabited by a wide range of microbes, or as he originally named them: *animalcules* (Leewenhoek 1684), but how the complex ecosystem of these organisms differs between healthy and diseased individuals has only recently come to the forefront. In the past, many approaches to understanding maladies of the oral cavity were reductionist in nature, relying on the classical and proven methods of microbiology as put forth originally by Koch (He and Shi, 2009). More recently, however, increased attention is being paid to understanding the oral cavity through microbial ecology by comparing healthy and diseased individuals, and this approach is beginning to increase our understanding of several diseases such as periodontitis and gingivitis (Wade 2013).

Recent work has even shown that not only does the community composition of patients with periodontitis differ significantly from healthy individuals, but also amongst diseased individuals, and that despite this, diseased individuals share increases in the same bacterial metabolic pathways as measured by metatranscriptomic analysis (Jorth et al. 2014). These results speak to the complex ecological nature of the oral microbiome, an environment that may be thought of as having smaller micro-niches in the form of functional opportunities that can be filled by a wide range of taxa depending

on many factors. But the oral cavity is just one amongst several human environments that have provided interesting results through the study of their microbial ecology.

Human skin, while contiguous across the body of most individuals, provides several unique microenvironments which microbial communities can exploit. These sites are often differentiated in their acidity, moisture level, sebum production, and invaginations (Grice and Segre 2011). Metataxonomic studies of the microbiome of human skin have shown that community structure is often correlated with common skin diseases such as psoriasis (Gao et al. 2008). In addition to the external portion of the skin, much of the internal portion of our bodies are also contiguous with the external environment, albeit with highly-differentiated membrane surfaces and conditions.

The vagina is one such environment, and has been shown to be highly unique in its microbial community (Lamont et al. 2011), which has been suggested to provide a protective environment against pathogenic and non-commensal organisms (Witkin et al. 2007). Other studies have attempted to attribute β-diversity between the vaginal microbiomes of ethnically-differentiated groups of women (Ravel et al. 2011, Ma et al. 2012), but these studies have failed to account for the effects of epigenetic factors that may be influenced by stress and environment. Perhaps one of the most interesting findings from the study of the vaginal microbiome is that infants born via caesarean section have a skin microbiome that differs from those born vaginally (Dominguez-Bello et al. 2010), and this difference can be partially mediated via exposure to vaginal fluids at birth (Dominguez-Bello et al. 2016).

The human gut microbiome

Of all the sites in and on the human body that harbor a unique microbial community, perhaps none has been more studied than the human gut. The gastrointestinal tract of humans is believed to harbor more than $7x10^{13}$ bacteria, roughly 70% of all the bacteria that constitute the entire human microbiome (Sekirov et al. 2010). Within the gastrointestinal tract, the local concentration of bacteria rises from the relatively low number of 10^1 cells in the stomach which is home to only a few taxa such as *Helicobacter pylori*, to the galactically-high number of 10^{12} cells in the distal colon, which is occupied by a much more diverse complement of bacteria (O'Hara and Shanahan 2006). This immense and complex environment is intricately linked to human biology, and indeed is even reflective of the evolutionary history of our and other species (Ley et al. 2008a; Ley et al. 2008b).

Long-term dietary and nutritional differentiation between mammals, including humans, is believed to have played an important role in molding the taxonomic groups that constitute our gut microbiomes (Muegge et al. 2010). Some authors have even reported the ability to elucidate phylogenetic relationships between apes by using metataxonomic profiles from their feces (Ochman et al. 2010), and while this trend may simply reflect dietary differences, there is an interesting relationship between humans and other great apes in relation to their respective gut microbiomes. It has been shown that, when compared to extant great apes, the gut microbiomes of modern humans represent a sharp decrease in α-diversity, a trend which has been correlated with so-called diseases of industrialization such as metabolic disorder (Moeller et al. 2014). The previously referenced study notwithstanding, most research into the human microbiome

(including the HMP) has centered around individuals living industrialized lifestyles, and this sampling bias has impeded our ability to ascertain the γ-diversity — or total diversity — of human gut microbiomes. Some researchers recognized the potential for this bias early on (Lewis et al. 2012), and called for a more diverse approach within and outside of the HMP.

Recent work from anthropologically-oriented research groups has started to remedy this sampling bias. A study from Schnorr et al. in 2014 compared the gut microbiomes of Hadza individuals from Tanzania to those of urban, industrialized Italians (Schnorr et al. 2014). The authors discovered enrichment amongst the Hadza for several genera, including *Treponema*, and were able to correlate their data with ethnographically collected lifestyle and dietary information. In doing so, the authors were not only able to postulate about potential lifeway variables driving the differentiation of the Hadza gut microbiome, but were also adding crucial data to the collection of human gut microbiomes previously published. Additional studies have also contributed valuable diversity to the global dataset. An article comparing the gut microbiomes of two Peruvian groups utilizing different subsistence strategies (the Matses, a hunter-gatherer group, and agriculturalists from Tunapuco) with those of members of an urban-industrialized North American city found not only strong separation between the microbial communities present in the North American and Peruvian groups, but also between the Peruvian groups based on their lifeways (Obregon-Tito et al. 2015). Additionally, these authors were able to recapitulate the findings from the Hadza study with the presence of *Treponema* in the Peruvian samples,

suggesting that this genus may be a widely spread member of the typical human gut microbiome that has been lost in industrialized peoples.

Considering the notion that diseases such as metabolic disorder are linked to a reduction in the diversity of the human gut microbiome, the discovery of genera such as *Treponema* in extant populations of humans living more traditional lifestyles may hint at what can be thought of as an ancestral microbiome. However, it is worth stating clearly that the populations from which these findings come, while ethnographically analogous to past populations, do not represent an unbiased representation of past traditional peoples. It is entirely possible that the human gut microbiome of people living agricultural or hunter-gatherer lifestyles has also undergone dramatic changes, albeit in potentially different directions with respect to the addition or loss of microbial taxa. In order to truly assess changes in the human gut microbiome over long periods of human history, we require the ability to sample from deep time points and screen for microbial DNA. Fortunately, our ability to do so has increased in the last 20 years to the point that recovering snapshots of evolution — be it human, microbial, or otherwise — is now possible.

Ancient DNA

The field of ancient DNA (aDNA) research has been around for nearly as long as we have had the capability to perform Sanger sequencing. The first published attempt by Russell Higuchi and colleagues to recover so-called fossil DNA from an extinct species of horse called the quagga is a now-famous citation (Higuchi et al. 1984), but perhaps more telling of the uphill battle the field would experience — though less famous — is

the comment by Alec Jeffreys accompanying the article, wherein he stated that the study of aDNA was 'nothing more than a glorious dream' (Jeffreys 1984). Of course the study of aDNA has turned out to be much more than a glorious dream, as the last thirty years' worth of research have shown (Hagelberg et al. 2015). As Hagelberg and colleagues point out, aDNA research has led to advances in forensics, conservation genetics, and our understanding of past human migrations. While the extraction and sequencing of aDNA may now seem routine, this ability has been earned through challenging experimentation and the fortuitous development that next-generation sequencing is well-suited to the nature of aDNA molecules.

Ancient DNA, unlike modern DNA, is characteristically modified in ways that present unique challenges to its extraction, purification, amplification by PCR and library construction, and subsequent sequencing. While this was not initially assumed to be the case (Pääbo 1985), several key findings have shaped our understanding about the nature of aDNA molecules. One of the characteristic traits of aDNA molecules comes from the analysis of modern DNA, where it was observed that the primary structure of this molecule — the phosphodiester bonds that constitute its "backbone" — is susceptible to spontaneous cleavage over time (Lindahl 1993). While this process is mediated in living tissues by the presence of the DNA repair machinery, no such mechanism is present in the environment of non-living tissues from which aDNA may be retrieved. For this reason, molecules of aDNA are characteristically and without exception fragmented to varying degrees (Sawyer et al. 2012); that is to say that aDNA molecules are *shorter* than DNA recovered from living tissues.

This reduction in fragment length has presented a unique challenge to the extraction of aDNA from archaeological or historical samples, as some methods designed for the extraction of DNA from modern samples are not sensitive to these shorter fragment lengths (Gaillard and Strauss 1990). Borrowing from improvements in molecular biology, in general, early aDNA scholars modified their extraction techniques to include purification via silica binding, first with in-solution silica particles (Höss and Pääbo 1993) and later by way of commercial silica-based spin columns (Yang et al. 1998). Continued improvements in the sensitivity and efficiency of recovering these shorter fragments has been an area of intense research within the field and has resulted in additional silica-based methods to improve DNA yields and purity (Rohland and Hofreiter 2007; Dabney et al. 2013a).

In addition the reduced fragment length of aDNA being a challenge for purification, it also poses a challenge for the direct amplification of these fragments via PCR in several ways. First, PCR relies on primers designed to target a locus of interest. Due to the largely stochastic processes by which DNA decays, these primer sites may have been cleaved in the aDNA molecules, preventing the primers from properly annealing and halting subsequent elongation. A related issue arises when contamination of less-fragmented (e.g. modern) DNA is present in the PCR reaction which shares the primer binding sites, as these molecules will be preferentially amplified, resulting in off-target PCR products (Handt et al. 1994). Finally, and owing to chemical modifications of aDNA which are discussed below, PCR amplification of aDNA can result in so-called jumping PCR, wherein the final products are much longer than the actual templates present, masking their authenticity as aDNA (Handt et al. 1994).

Fortunately, the advent of next-generation shotgun sequencing has provided a workaround for these issues. Direct amplification of aDNA molecules via PCR is often eschewed in favor of first building DNA libraries consisting of the original fragments to which known primer-binding sequences are ligated. This library construction allows for amplification of each DNA fragment in a sample using a single set of primers, bypassing most of the laboratory issues associated with fragmented aDNA.

Fragmentation, however, is not the only way in which aDNA molecules differ from modern ones. It was discovered quite early on that, in addition to their shorter length, aDNA molecules contained chemical modifications to their nucleic acid residues (Höss et al. 1996). Of these chemical modifications, the most abundantly observed is the hydrolytic deamination of cytosine residues to uracil (Hofreiter et al. 2001; Brotherton et al. 2007). This deamination of cytosine to uracil results in characteristic patterns of DNA damage, particularly at single-stranded 5' overhangs of fragmented molecules (Briggs et al. 2007).

Due to the way in which DNA polymerases function during PCR, uracil residues resultant from cytosine deamination will be encoded as thymine (C to T) instead of guanine on complementary strands, and subsequent complementary strands of this template will contain an adenine instead of a guanine (G to A) (Hofreiter et al. 2001). These so-called miscoding lesions present themselves after aligning sequenced aDNA reads to reference genomes in the form of alignment mismatches at the 5' and 3' termini of reads. While this deamination can be corrected using uracil-DNA-glycosylase (Briggs et al. 2010; Rohland et al. 2014), the presence of these miscoding lesions has been employed in the software package mapDamage as a verification

method for aDNA studies, as modern DNA should not exhibit these characteristic damage patterns (Ginolhac et al. 2011, Jónsson et al. 2013). Here again is the field of aDNA research fortuitously supported by NGS methods, as it is the high number of sequencing reads generated via these methods that permit the type of statistical analyses performed by applications such as mapDamage.

Despite patterns of fragmentation and deamination being ubiquitous in DNA recovered from ancient samples, much is still unknown about the processes which result in them. While some have concluded that DNA behaves in a manner concordant with exponential decay (Allentoft et al. 2012), more recent work has shown that fragmentation occurs first very rapidly before leveling off, while deamination is in agreement with an exponential decay model (Kistler et al. 2017). As more and more aDNA datasets become available, our ability to understand the kinetics of this molecule will improve (Dabney et al. 2013b). In the meantime, our current understanding of these phenomena, coupled with best practices to avoid modern-day contamination (Cooper and Poinar 2000), have permitted incredible findings which previously seemed out of our reach. Studies of aDNA have yielded complete genomes from Neanderthals (Green et al. 2010) and previously unknown hominin species (Reich et al. 2010). Additionally, aDNA analysis of anatomically modern human samples is constantly refining our understanding of human dispersals and migrations (Allentoft et al. 2015, Laziridis et al. 2014). However, it is not only humans and other animals that have left traces of their genetic selves for us to study, but also microorganisms.

Microbial Archaeology

The advent and growth of the field of ancient DNA in archaeology and forensics was joined by interests from other fields, including microbiology. Mirroring the rest of aDNA research, many early reports of ancient microbial DNA included lofty claims of hyper-old DNA, sourced from ice core or geological samples which were many millions of years old (Willerslev et al. 2004), with one team of researchers going so far as to say they had recovered viable bacterial endospores encased in amber that dated back 25 million years (Cano and Borucki 1995). While these early findings remain a topic of debate, subsequent work in recovering aDNA from bacteria and archaea has been important in understanding their evolution and preservation within various substrates and has important implications for the field of astrobiology (Gilichinsky et al. 2007, Sankaranarayanan et al. 2014). Beyond environmental studies of microbial aDNA, however, a new field is emerging in the form of *microbial archaeology*.

It has been said that humans are simply "animals in a bacterial world", as evidenced not only by our microbiomes and the fact that bacteria inhabit essentially every niche on Earth, but also that all life on Earth owes its very existence to these unseen organisms (McFall-Ngai et al. 2013). As such, humans have never existed apart from their microbial counterparts — be they commensal or pathogenic. The pathogenic cohort of bacteria that affect humans have long been of interest to archaeologists, and there is a sub-discipline of archaeology concerned with studying the diseases they cause: paleopathology. This science, situated between archaeology and physical anthropology, has been critical in understanding infectious disease in past peoples, but has classically relied upon skeletal analysis to make its conclusions.

Unfortunately, skeletal analysis alone is insufficient for absolute diagnosis of some infectious diseases, such as syphilis, which is caused by the bacterium *Treponema pallidum* (Harper et al. 2011). In the case of other diseases such as tuberculosis, whose identification via skeletal lesions is an accepted practice, a lack of scientific rigor has often led to misdiagnosis (Wilbur et al. 2009). Therefore, some paleopathologists have turned to biomolecular analysis as a means to verify inferences made from skeletal analysis and as a way to study the spread and evolution of disease-causing bacteria. These biomolecular analyses have taken several forms, but mostly center around the genetic identification of pathogens through aDNA sequencing. While the earliest approaches attempted to target known loci of pathogenic bacteria via PCR (Salo et al. 1994; Zink et al. 2001), advances in NGS technology and target sequence enrichment (Schuenemann et al. 2011) are allowing biomolecular paleopathologists to reconstruct entire genomes of pathogens from archaeological samples (Bos et al. 2011, Bos et al. 2014).

Ancient microbiomes

In addition to the study of individual bacterial species from archaeological samples, there is growing interest in extending microbiome science into the distant past to better understand past diet, nutrition, and human health (Warinner et al. 2015). The study of ancient microbiomes presents its own unique set of challenges, not least of which is the post-mortem shift of microbial communities. As stated earlier, the human microbiome is deeply integrated with human physiology, and once that environment (i.e. a living human) ceases to exist, the microbial community that relies on it will

experience changeover in accordance with decomposition (Metcalf et al. 2013, Metcalf et al. 2016). This means that, except in exceedingly rare circumstances, the endogenous microbial ecology of external or nearly external portions of the human body — such as the skin, urogenital tract, or nasal cavity — will be replaced by environmental taxa, rendering studies of the original microbiomes impossible. There are, however, archaeological remains which do maintain a biological signature of the oral and gut microbiomes as they existed during a person's life: fossilized dental plaque, known as calculus, and palaeofaeces or *coprolites,* respectively.

During an individual's life, a calcified biofilm known as dental plaque forms on the surfaces of the teeth through an organized process of bacterial colonization (Rosan and Lamont 2000). The taxa that constitute this biofilm number in the hundreds (Paster et al. 2001), and their role in oral health and disease remains a lively area of research (Peterson et al. 2013, Krishnan et al. 2016). Due to the calcified nature of dental plaque, it represents the only tissue of the human body that can be said to fossilize during an individual's lifetime (Warinner et al. 2015), and thus is persistent in the archaeological record in a form known as dental calculus in populations who did not practice mechanical removal of this material. In recent years, dental calculus has been recognized as one of the richest sources of ancient biomolecules including proteins (Warinner et al. 2014a), human DNA (Ozga et al. 2016), and microbial DNA (Warinner et al. 2014b, Weyreich et al. 2014). Thus, dental calculus has been shown to be capable of providing a snapshot of the oral microbiome at different points in human evolutionary history. Prominent physical anthropologists have already weighed in on the importance of understanding the evolution of the oral microbiome in relation to

dietary and lifeway changes (Harper and Armelagos 2013), and early work has demonstrated how these lifeway changes have impacted the microbial ecology of the oral cavity, particularly the shift towards cariogenic taxa during the Industrial Revolution (Adler et al. 2013).

As stated earlier, the oral microbiome is not the only human-associated microbial environment which we may recover from archaeological remains. First, it is important to note that in modern studies of the human gut microbiome, deposited feces is used as a proxy for the microbial community present in the distal colon (Peterson et al. 2009). Coprolites, while not freshly deposited, are the archaeological equivalent to the samples used in modern studies of the gut microbiome. Coprolites are formed when the feces of past peoples (or other animals) are preserved through either partial mineralization or desiccation (Reinhard and Bryant 1992). These materials have been recovered from caves (Jenkins et al. 2012), latrines (Reinhard 1986), and *in situ* within the intestines of human remains (Luciani et al. 2006), and are a rich source of dietary and parasitological information. Additionally, coprolites have been used a source material for endogenous aDNA from the species which deposited the feces (Poinar et al. 1999; Bon et al. 2012).

Most recently, coprolites have become of great interest to researchers seeking to understand the evolution of the gut microbiome of humans. The first studies to perform such analysis using NGS techniques successfully reconstructed microbial communities closely resembling the microbiome of the modern human gut using both shotgun (Tito et. al 2008) and 16S rRNA gene amplicon sequencing (Tito et al. 2012), despite the post-depositional taxonomic changes expected for ancient samples. These findings

suggest that, under preferable taphonomic conditions, coprolites are a suitable substrate for exploring the ancient human gut microbiome.

Subsequent work in the study of ancient oral microbiomes has shown that a shotgun sequencing approach is favorable to 16S rRNA gene amplicon sequencing for reconstructing microbial communities due in part to the fragmented nature of aDNA molecules (Ziesemer et al. 2015). Beyond sequencing strategy, however, little work has been published comparing methods for the extraction and purification of aDNA from coprolites. Tito et al. did compare a commercial soil DNA isolation kit with an isopropanol precipitation protocol in their 2008 article (Tito et al. 2008), however, different samples were not treated with both protocols limiting their ability for direct comparison. Recent advances in the isolation and purification of aDNA molecules warrant a thorough comparison of available methods to determine best practices for working with coprolites, a scientifically valuable and scarce archaeological resource.

Chapter 2: Coprolites

Coprolite studies to date

The study of coprolites has historically focused on palynological and parasitological analyses through flotation and microscopy, but molecular analyses including aDNA extraction and sequencing have become more common. One of the first published reports of aDNA recovery from a coprolite was in 1998 (Poinar et al. 1998), wherein PCR-amplified mitochondrial 12S rRNA gene fragments were used to identify the organism which deposited the feces as the extinct ground sloth *Nothrotheriops shastensis*. Subsequent experiments on coprolites have recovered

endogenous nuclear DNA (Poinar et al. 2003), as well as DNA from parasites (Loreille et al. 2001; Iñiguez et al. 2006), viruses (Appelt et al. 2014; Ng et al. 2014), dietary sources (Wood et al. 2008; Bon et al. 2012), and the microbiota of the gastrointestinal tract (Obregon-Tito et al. 2008; Obregon-Tito et al. 2012).

While all these studies have shown that coprolites can provide a rich source of various genetic information, they have done so using a wide variety of DNA extraction and purification techniques, and none have systematically tested the effects of laboratory procedures on the ability to recover aDNA from coprolites, and in the case of ancient microbiome studies, how laboratory procedures impact downstream metataxonomic analyses. Several procedural factors may impact the recovery and purification of aDNA from coprolites, including cell lysis and DNA purification strategies. The various methods for cell lysis and DNA purification strategy which have been used in coprolite studies to date are discussed in the following section.

Lysis and DNA purification strategies with coprolites

Coprolites are typically found in arid locations which allow for the rapid desiccation of feces, and this desiccation leads to the preservation of intact cells as well as whole macro and microorganisms (Reinhard et al. 1986). As DNA is presumably contained within these cell, lysis is commonly performed prior to DNA purification in order to increase the amount of DNA molecules recoverd. One method employed in previous coprolite studies is to chemically lyse cell membranes prior to DNA purification via the addition of a lysis buffer containing cell membrane permeability enhancers such as sodium dodecyl sulphate (SDS) or other detergents. Studies which

have utilized commercial kits designed for extraction of DNA from modern fecal samples have also incorporated a mechanical lysis step in the form of bead beating.

Following the liberation of DNA from intact cells, it is purified to remove proteins and other off-target biomolecules, as well as inhibitory compounds. DNA purification in the context of coprolite studies has been performed via several different methods. Early experiments such as the one performed by Poinar and colleagues (Poinar et al. 1998) utilized a standard phenol-chloroform extraction, followed by concentration of the DNA in a microconcentrator to avoid losses associated with ethanol precipitation (Pääbo et al. 1988), while other studies have disregarded the issues associated with ethanol precipitation and (Cano et al. 2000). More recently, the purification and concentration of DNA recovered from coprolites has been accomplished with silica based spin columns such as those provided in commercial kits designed for the recovery of DNA from soil or feces (Appelt et al. 2014; Tito et al. 2008; Tito et al. 2012).

The unknowns of aDNA from coprolites

The effects of cell lysis and DNA purification strategies when working with coprolites merit systematic testing for several reasons. First, it is important to identify protocols which routinely recover the greatest amount of DNA from coprolites. This is especially critical when the goal of the study is genetic identification through endogenous DNA, as it has been shown that the endogenous DNA content of fresh fecal samples is vastly outnumbered by bacterial DNA, and requires sequence enrichment strategies such as targeted DNA capture to obtain sufficient genome coverage (Perry et al. 2010). This

effect is enhanced by the degraded and fragmentary nature of aDNA in archaeological samples, and has resulted in specialized capture methods for the recovery of low abundance genomic fragments from ancient bone and tooth samples (Carpenter et al. 2013). While sequence enrichment strategies such as whole-genome capture improve genomic coverage from feces and archaeological samples, they are not perfectly efficient, and thus benefit from having the greatest number of starting molecules from which targets may be enriched, supporting the need for extraction and purification methods which maximize initial concentrations of DNA.

In addition to testing different DNA extraction and purification methods for their ability to maximize the recovery of aDNA from coprolites, further experiments are also warranted that systematically test whether laboratory procedures impact the observed community structure in studies of ancient human gut microbiomes as derived from coprolites. Experiments performed with fresh fecal samples have demonstrated that choice of extraction and purification protocols can significantly impact the downstream assignment of sequencing reads to taxonomic groups (Wesolowska-Andersen et al. 2014). At present it is unknown if microbiomes reconstructed from coprolites are also affected in this manner, as the taxonomic shifts of coprolites post-deposition are not well characterized.

The present study

The present study was designed to systematically compare DNA yields and reconstructed microbiomes as generated by five different DNA extraction protocols, including the protocol used by the HMP, a modified version of that protocol which uses

a silica-based spin column purification as described in *Dabney et al. 2013*, and three methods designed specifically for the recovery of aDNA which are all described in more detail in Chapter 3. The study was conducted under the hypothesis that methods designed for the recovery of aDNA would yield higher DNA concentrations, and therefore allow for more accurate reconstruction of ancient gut microbiomes. The coprolites used in this study have been previously shown to contain intact aDNA sufficient for microbiome studies (Tito et al. 2008; Tito et al. 2012). These coprolites were recovered from La Cueva de los Muertos Chiquitos, an archaeological site associated with the Loma San Gabriel culture that occupied present-day Durango, Mexico approximately 1,400 years before present.

Chapter 3: Materials and methods

Methods

Five DNA extraction methods were tested using three human coprolites (*Zape 2, Zape 5, Zape 28*) from the Rio Zape archaeological site in Durango, Mexico. The extraction methods included standard and modified protocols for fecal DNA isolation employed by the Human Microbiome Project (HMP) and three protocols adapted from Dabney et al. (2013) that are designed for the recovery highly-degraded ancient DNA (aDNA). All sample preparation and DNA extraction procedures were carried out at the Laboratories of Molecular Anthropology and Microbiome Reseaarch (LMAMR) at the University of Oklahoma in an ISO-6 cleanroom facility dedicated to the extraction of ancient biomolecules. Laboratory protcols were carried out in accordance with established contamination control precautions and workflows, and for each extraction

method, a non-template extraction control (negative control) was processed alongside the experimental samples during all analytical steps to monitor for possible contamination.

Extraction methods

Extraction method 1: HMP protocol

DNA was extracted from coprolite material using the MoBio PowerSoil kit according to manufacturer instructions. In brief, approximately 200 mg of sample was added to a MoBio PowerBead tube with 750 µl of MoBio Bead Solution. Samples were then rotated for 2 hours. A volume of 60 µl of MoBio Solution C1 was added and the samples were vortexed briefly before being incubated at 65°C for 10 minutes. Samples were then subjected to bead beating for 10 minutes before being centrifuged at 13,000 rcf for 1 minute. The supernatant was then transferred to a new microcentrifuge tube, into which 250 µl of MoBio Solution C2 was added. Samples were then vortexed briefly before incubation at 4° C for 5 minutes, followed by centrifugation at 13,000 rcf for 1 minute. The supernatant was then transferred to a new microcentrifuge tube, to which 200 µl of MoBio Solution C3 was added. Samples were then vortexed briefly and incubated at 4° C for 5 minutes, followed by centrifugation at 13,000 rcf for 1 minute. The supernatant was again transferred to a new microcentrifuge tube, and 1200 µl of MoBio Solution C4 was added. Samples were loaded onto a MoBio spin filter in approximately 700 µl increments ($n=3$), centrifuging the spin filter at 13,000 rcf for 1 minute each time to bind DNA to the column. Next, 500 µl of MoBio Solution C5 was then added to the spin filter and centrifuged at 13,000 rcf for 1 minute. The effluent was

discarded, and the spin filter was again centrifuged at 13,000 rcf for 1 minute to dry the column. Spin filters were then transferred to new low-bind microcentrifuge tubes, and 60 µl of MoBio Solution C6 was added. Samples were then eluted from the spin filters via centrifugation at 13,000 rcf for 1 minute. The total volume of extracted DNA was 60 µl.

Extraction method 2: HMP protocol with modified MinElute protocol

This method was intended to compare the differential recovery of short DNA fragments between silica column protocols. The MoBio spin column is optimized for retaining long DNA fragments, whereas the Qiagen MinElute column in tandem with the modifications from Dabney et al. (2013) is designed to maximize the recovery of short fragments. This method was identical to Extraction Method 1 through the step in which samples were treated with MoBio Solution C3, incubated at 4°C for 5 minutes, and then centrifuged at 13,000 rcf for 1 minute. Following this centrifugation, DNA was purified and concentrated according to *Dabney et al. 2013*. Briefly, the supernatant and 14 mL of Qiagen PB buffer were loaded into Qiagen MinElute columns that were fitted to Zymo Spin V reservoirs and placed into a 50 mL falcon tube. Samples were centrifuged at 1,500 rcf for 4 minutes, rotated 90° in the centrifuge, and then spun at 1,500 rcf for an additional 2 minutes. The MinElute columns were then removed from the Zymo reservoirs and transferred to microcentrifuge tubes. Samples were washed twice with 700 µl of Qiagen PE buffer by centrifugation at 10,000 rcf for 1 minute. The effluent was discarded, and the MinElute columns were transferred to new microcentrifuge tubes. A volume of 30 µl of Qiagen EB buffer was added to the

columns and incubated at room temperature for 5 minutes. Samples were then centrifuged at 13,000 rcf for 1 minute to elute DNA. An additional 30 µl of Qiagen EB buffer was added to the column, incubated at room temperature for 5 minutes, and DNA was again eluted into the same collection tube via centrifugation at 13,000 rcf for 1 minute. The total volume of extracted DNA was 60 µl.

Extraction method 3: Phenol-chloroform with modified MinElute protocol

This extraction method was designed to determine whether a phenol-chloroform extraction improved DNA yields and subsequent library preparation through removal of off-target biomolecules and PCR inhibitors. Approximately 200 mg of sample was added to a MoBio PowerBead tube with 750 µl of MoBio Bead Solution. 400 µl of 0.5M EDTA and 100 µl of Proteinase K were added to each tube. Samples were then rotated for at room temperature for 4 hours. Samples were then centrifuged at 6,000 rcf for 5 minutes, and the supernatant was removed and added to a new microcentrifuge tube containing 750 µl of a 25:24:1 mixture of phenol:chloroform:isoamyl alcohol. Samples were mixed by gentle vortexing and then centrifuged at 13,000 rcf for 1 minute. The aqueous phase was transferred to a new microcentrifuge tube containing an additional 750 µl of a 25:24:1 mixture of phenol:chloroform:isoamyl alcohol. Samples were mixed by gentle vortexing and again centrifuged at 13,000 rcf for 1 minute. The aqueous phase was transferred to a new microcentrifuge tube containing 750 µl of a 24:1 chloroform:isoamyl alcohol mixture. Samples were mixed by gentle vortexing and and centrifuged at 13,000 rcf for 1 minute. Following this centrifugation, DNA was purified and concentrated according to the protocol described in Extraction Method 2.

The total volume of extracted DNA was 60 µl.

Extraction method 4: Modified MinElute column

This method is adapted from Dabney et al. (2013), with the addition of a bead-beating step to homogenize the sample. Approximately 200 mg of sample was added to a MoBio PowerBead tube with 750 µl of MoBio Bead Solution. A volume of 400 µl of 0.5M EDTA and 100 µl of Proteinase K were added to each tube. Samples were then rotated for at room temperature for 4 hours. Samples were then centrifuged at 6,000 rcf for 5 minutes. Following this centrifugation, DNA was purified and concentrated according to the protocol described in Extraction method 2. The total volume of extracted DNA was 60 µl.

Extraction method 5: Double modified MinElute column

The purpose of this method was to test a solution to a problem encountered in earlier experiments wherein MinElute columns would become clogged. It was identical to Extraction Method 4, except the supernatant from each sample following the centrifugation at 6,000 rcf for 5 minutes was divided into two equal aliquots, each of which were then carried through the remaining steps of Extraction Method 4. Additionally, the final elution of DNA with Qiagen EB buffer was reduced to two rounds of 15 µl, and the elutions were then pooled for a total of 60 µl. The total volume of extracted DNA was 60 µl.

DNA extract quantification

The DNA concentration of all extracts was quantified using a Qubit 3.0 Benchtop Fluorometer using the Qubit High Sensitivity dsDNA reagents according to the manufacturer's instructions with 2 µl of extract. DNA yields were normalized to the amount of starting material used by the following formula:

$$\frac{concentration * elution\ volume}{weight\ of\ starting\ material}$$

DNA library construction

Approximately 100 ng of DNA as calculated from Qubit quantitation was added to molecular grade water in a clean microcentrifuge tube for a total volume of 42.5 µl. A volume of 5 µl of NEBNext End Repair Reaction Buffer (10X) and 2.5 µl of End Repair Enzyme Mix were added to each sample. Samples were gently mixed by hand and incubated at room temperature for 40 minutes, followed by an additional incubation at 37°C for 15 minutes. Samples were then loaded into Qiagen MinElute columns with 250 µl of Qiagen PB buffer. The columns were centrifuged at 6,000 rcf for 1 minute, and the effluent was discarded. 750 µl of Qiagen PE buffer was added to the columns, and they were centrifuged at 10,000 rcf for 1 minute. The effluent was discarded, and the columns were centrifuged at 13,000 rcf for 1 minute to dry. Columns were transferred to new microcentrifuge tubes, 30 µl of Qiagen EB buffer was, and samples were incubated at 37°C for 15 minutes. Following incubation, the end-repaired DNA was eluted from the columns by centrifugation at 13,000 rcf for 1 minute.

A volume of 10 µl of Quick Ligation Reaction Buffer (5X), 3 µl of an adaptor mix containing the Illumina P5 and P7 adaptors (µM), 2 µl of molecular grade water,

and 5 µl of Quick T4 Ligase were added to the 30 µl of end-repaired DNA. Samples were mixed gently by hand and incubated at room temperature for 20 minutes. Samples were then loaded into Qiagen MinElute columns with 250 µl of Qiagen PB buffer. The columns were centrifuged at 6,000 rcf for 1 minute, and the effluent was discarded. 750 µl of Qiagen PE buffer was added to the columns, and they were centrifuged at 10,000 rcf for 1 minute. The effluent was discarded, and the columns were centrifuged at 13,000 rcf for 1 minute to dry. Columns were transferred to new microcentrifuge tubes, 30 µl of Qiagen EB buffer was, and samples were incubated at 37°C for 15 minutes. Following incubation, the adaptor-ligated DNA was eluted from the columns by centrifugation at 13,000 rcf for 1 minute.

Thirteen µl of water, 5 µl of Adapter Fill-In Reaction Buffer, and 2 µl of *Bst* DNA polymerase were then added to the 30 µl of adapter-ligated DNA extracts. Samples were mixed gently by hand and incubated at 37°C for 30 minutes, followed by incubation at 80°C for 20 minutes to inactivate the enzyme. Samples were then placed in a -20°C freezer overnight. Samples were then thawed and prepared for a test qPCR using the following reagents (per extract): 1 µl of sample, 12.5 µl of KAPA HiFi HotStart Uracil Ready Mix (2X), 5 µl of water, 1 µl BSA (2.5 mg/ml), 0.75 µl of 10 µM indexed i5 primer, 0.75 µl of 10 µM indexed i7 primer, and 4 µl of 50 µM SYTO 9. The qPCR reaction was then run on a Roche LightCycler 96 with the following program: initial denaturing at 98°C for 30s, 45 cycles (denaturing at 98°C for 10s, annealing at 60°C for 15s, elongation at 72°C for 30s), and a final elongation at 72°C for 5 minutes.

The results of the test qPCR reaction were used to determine the number of amplification cycles in the subsequent indexing step. All extracts were then prepared in

triplicate for an indexing PCR using unique i5 and i7 Illumina sequencing primers with the following reagents (per extract): 4 µl sample, 12.5 µl KAPA HiFi HotStart Uracil Ready Mix (2X), 6 µl water, 1 µl BSA (2.5mg/ml), 0.75 µl each of a 10 µM unique i5 and i7 primer. Samples were then amplified according to the number of cycles determined by the test qPCR on AnalytikJena thermocyclers with the following program: initial denaturing at 98°C for 30s, followed by various cycles of: denaturing at 98°C for 10s, annealing at 60°C for 15s, elongation at 72°C for 30s, and a final elongation at 72°C for 5 minutes.

The amplified DNA libraries were then checked for successful amplification by gel electrophoresis on a 2% agarose gel. The triplicate reactions were then pooled into single tubes and purified using Qiagen MinElute columns. The size distribution of each library was then calculated with an Agilent BioAnalyzer using the High Sensitivity DNA kit. Samples were then pooled in equimolar amounts and cleaned of residual primers and adapter dimers via PippinPrep. The completed libraries were then sent for 2x100 paired-end Illumina sequencing at the Yale Center for Genome Analysis.

Shotgun data analysis and quality filtering

All reads were initially screened for sequencing quality with FastQC. Reads were then merged and sequencing adapters were removed using AdapterRemoval (Schubert, Lindgreen, and Orlando 2016) with the following settings: --maxns 0 --trimqualities --minquality 30 --collapse --minlength 25 --minalignmentlength 10. Subsequent removal of remaining adapter contamination was performed by mapping all reads to the adapter sequences using the bowtie2 aligner and the script filter_fasta.py

from the QIIME software package (Caporaso et al. 2010). The resulting reads were then used for all further analyses.

Metataxonomic characterization

For metataxonomic analysis using QIIME, all analysis-ready reads were first mapped to the greengenes (v13.8) 16S rRNA gene database (DeSantis et al. 2006) using the bowtie2 aligner (Langmead and Salzberg 2012). All mapped reads were then sorted and de-duplicated using SAMTools (Li et al. 2009) and a FASTA file containing the resulting reads was generated using the QIIME script filter_fasta.py and the FASTX-Toolkit (Gordon and Hannon 2010). Sequence headers were formatted sequentially for QIIME analysis using a simple awk command. Operational Taxonomic Unit (OTU) clustering was performed against the greengenes 16S rRNA gene database (version 13.8) preclustered at 97 % sequence similarity using QIIME's pick_closed_reference_otus.py script with the following parameters: --max_accepts 20, --max_rejects 500, --word_length 12, --stepwords 20, --enable_rev_strand_match True. The resulting OTU file was converted into a BIOM formatted table (McDonald et al. 2012) which was used to generate OTU rarefaction data via QIIME's parallel_multiple_rarefactions.py script.

While the rarefaction curves did not display the characteristic leveling for a sample whose community has been sampled to a sufficient depth, this is expected for non-amplicon sequencing as shown in previous studies (Ranjan et al. 2015). At a depth of 7,000 reads, clear separation was obtained between the samples and the negative controls, and this value was used to rarify the data for further analyses. Alpha and beta

diversity were computed using QIIME's built-in scripts. Relative abundance of taxa at the genus level (species where available) was calculated using the summarize_taxa.py script in QIIME. To overcome the potential limitations of 16S rRNA gene based OTU binning in a shotgun sequencing dataset, the Metagenomic Intra-Species Diversity Analysis System (MIDAS) was also used to estimate species abundance (Nayfach et al. 2016). Merged analysis-ready reads were used as the input for the run_midas.py script using the species operator with the default settings. The output of the resulting species abundance analyses were merged using tools within MIDAS, and the subsequent species abundance table was then used for comparison with the output of QIIME.

Bayesian source tracking of microbial communities

SourceTracker (Knights et al. 2011) was used to predict the source of the microbial communities present in each sample. OTUs generated by QIIME were merged with an OTU table containing source communities from soil (Johnson et al. 2016), the gut of urban humans (Sankaranarayanan, 2015), the gut of rural humans (Obregon-Tito et al. 2015, Schnorr et al. 2013), human skin (Oh et al. 2016), and sub- and supra-gingival plaque (Peterson et al. 2012). This combined OTU table was converted to the BIOM format, and was used in conjunction with a mapping file containing the aforementioned source communities as sources and the samples from the present study as sinks as input for SourceTracker with the default settings.

Ancient DNA validation, insert length analysis, and statistical testing

The taxa with the highest relative abundance from the QIIME and MIDAS analyses were compared to identify those shared between the pipelines. From these, the top 21 taxa were selected, and reference genomes were downloaded from NCBI. All reads for each sample were then mapped to each of these reference genomes using the bowtie2 aligner. The resulting SAM files and the reference genomes were then used as input for the application mapDamage 2.0 to calculate aDNA damage patterns. Additionally, insert lengths were computed for all reads against each reference genome using a custom Python script. The effective library size for each sample was computed independent of mapping by sorting the collapsed reads from each sample by size, and then clustering within each sample by 99% identity using vsearch (Rognes et al. 2016). Statistical tests to elucidate relationships between DNA damage patterns, insert lengths, effective library size, and other metadata were then conducted in R.

Results

DNA extraction yields

DNA yields from extractions that were purified using the Qiagen MinElute column ($n=24$) significantly ($p<0.0001$) outperformed those which used the MoBio column ($n=6$) as determined by a Mann-Whitney-Wilcoxon Test, and this difference was independent of the experimental replicates "A" and "B" (Figure 1). A complete table of the normalized DNA yields is also provided (Table 1).

Reads

To determine the quality of the sequencing data and prepare analysis-ready reads, all files were trimmed, merged, and quality-filtered with AdapterRemoval. The mean number of reads generated for each sample, excluding negative controls, was 14,594,955 prior to quality filtering (*minimum=7,011, median=11,056,906, maximum=59,880,666*). The mean number of paired-end reads which were successfully merged was 13,047,463 (*minimum=3359, median=9,865,675, maximum=53,378,272*), representing an average of 87.78%. The mean number of paired-end reads which were not successfully merged was 209,942 (*minimum=185, median=106,616, maximum=1,859,799*), representing an average of 1.50%. Due to the low number of reads (*n = 7,011*) obtained for the Zape 2B sample prepared with Extraction Method 4, it was omitted from downstream analyses. A summary of all reads pre- and post-quality filtering is available in Table 2.

Metataxonomic characterization

The QIIME pipeline was used to characterize the microbial community of each sample. The number of reads that mapped to the 16S rRNA gene in the greengenes database and was subsequently used for QIIME analysis is displayed in Table 2. Following OTU clustering and rarefaction of the data, a weighted UNIFRAC beta-diversity calculation was performed. This analysis revealed distinct community structure between samples, but importantly not between the extraction methods or purification column type (Figure 2).

To corroborate the metagenomic communities discovered through analysis with QIIME, the MIDAS pipeline was used to estimate the abundance of microbial

community members from each sample based on multiple phylogenetically informative marker genes such as those coding for protein synthesis, cell division, and DNA repair. The result of this analysis revealed several taxa that were shared between MIDAS and QIIME, suggesting that that the coverage for the 16S rRNA gene in the dataset was sufficient for accurate ancient metagenome reconstruction. From these taxa, 21 genera (or species where available) were selected that had at least 0.05 relative abundance across all samples for downstream analysis. A summary of these taxa is available in Table 3. Notably, in both QIIME and MIDAS, the common intestinal bacteria *Prevotella copri* was identified as the most abundant taxon.

Bayesian source tracking

To verify that the origin of the microbial DNA present in each sample was the human gut, Bayesian source tracking was performed with SourceTracker. The results of this analysis showed that the microbial communities present at the OTU level were constituted largely by OTUs present in the human gut, particularly those OTUs belonging to the rural human gut as measured by Obregon-Tito, 2015 and Schnorr, 2013. Additionally, none of the samples showed significant contributions to their makeup from common sources of contamination such as human skin, the human oral cavity, or soil.

These results support that the coprolites being studied are sufficient proxies for the gut microbiomes of the individuals from which they originated and do not reflect significant post-deposit microbial community alteration from the environment. The predictions of source communities for each sample are displayed in Figure 3. A

weighted UNIFRAC β-diversity analysis of the OTU table containing the sources and the samples from the present study was performed to further confirm these results. All of the samples from the present study fall within the range of diversity of the human gut samples that were used as sources as visualized in the principal component analysis plot in Figure 4. This result also recapitulates the finding stated above, that all extraction methods produce data which will subsequently produce microbial communities that are comparable to one another.

Ancient DNA analysis

To verify that the data generated were from authentic aDNA, all reads which were mapped to the 21 reference bacterial genomes were processed through the application mapDamage 2.0. When grouping samples by extraction method, most showed damage patterns at the 5' and 3' termini of inserts consistent with aDNA. However, the median percent of damage at the first base position in the phenol:chloroform extraction method was significantly lower than the other methods as determined by a Kruskal-Wallis H test. The distributions of these damage patterns within each sample and method is presented in Figure 5.

Fragment length analysis

To compute the lengths of the recovered aDNA fragments, a custom Python script was used to extract mapping coordinates from the SAM files generated by mapping each sample to the 21 reference bacterial genomes. The results of this analysis display distributions of insert lengths for all samples and extraction methods consistent with expectations for aDNA (Figure 6). Statistical analyses were then performed in R to

determine if DNA damage or insert length were driven by experimental variables. A Mann-Whitney-Wilcoxon Test showed that there was a highly significant ($p < 2.2e-16$) increase in the median insert length in the methods which utilized the modified MinElute column protocol for DNA purification (Figure 7), suggesting that while being modified to increase the recovery of shorter fragments, the modified MinElute extraction method outperforms the MoBio column for molecules of all sizes. This finding supports the overall higher recovery of DNA using the MinElute column as shown previously in Figure 1.

Examining relationships between DNA damage, insert length, and experimental variables

Linear modeling of insert lengths and DNA damage with other experimental variables was performed to further explore the data generated. Because *P. copri* had the highest overall abundance in QIIME and MIDAS, as well as the greatest amount of coverage in MIDAS, data for this bacterium were used for these tests. These analyses revealed several interesting results. First, insert length was shown to be positively correlated with the normalized amount of DNA extracted from each sample, however the data do not strongly adhere to the linear regression ($R^2 = 0.54026$). Secondly, DNA damage patterns show an inverse relationship to that of the insert length, and are negatively correlated with DNA yields, albeit with even greater deviation from the linear model ($R^2 = 0.25397$). To verify that these findings were not a result of sequencing bias, insert lengths and DNA damage pattern frequency were each modeled against the effective size of each sequencing library.

The results of these analyses showed that neither insert length nor DNA damage patterns could be predicted by effective library size. A summary of these linear models is available in Figure 8. To determine if there were any taxonomic influences on the observed levels of DNA damage and insert length, z-scores were calculated using the mean value and standard deviation of each metric. The results of this analysis indicate that the proportion of cytosine deamination and the length of inserts are both not due to differential preservation for the 21 taxa to which reads were mapped (Figures 9 and 10).

	HMP		HMP + MinElute		P:C MinElute		MinElute		2X MinElute	
	A	B	A	B	A	B	A	B	A	B
Zape 2										
ng DNA/mg Material	1.01	0.75	6.42	2.40	9.89	6.98	7.05	4.84	9.20	11.70
Zape 5										
ng DNA/mg Material	0.04	0.02	0.10	0.22	3.21	2.67	6.71	2.98	4.22	4.80
Zape 28										
ng DNA/mg Material	0.19	0.03	2.63	0.87	5.41	7.15	2.87	2.02	2.28	5.08

Table 1. Normalized DNA yields

Method	Sample	Total Reads	% Collapsed	% Paired	% to Qiime
HMP Protocol	Zape 02	65806863	89.6	0.85	0.29
	Zape 05	17936998	90.9	0.6	0.19
	Zape 28	31216395	91.8	1.02	0.25
HMP + MinElute	Zape 02	28416026	88.1	1.73	0.41
	Zape 05	39180651	94.2	0.51	0.18
	Zape 28	23697831	93	0.5	0.27
P:C MinElute	Zape 02	51284345	85.1	5.21	0.39
	Zape 05	14815130	87.3	1.61	0.21
	Zape 28	39429925	90.6	0.61	0.31
MinElute	Zape 02	17801482	65.1	2.9	0.35
	Zape 05	22113811	87.4	1.78	0.21
	Zape 28	34750101	93.2	0.52	0.26
2X MinElute	Zape 02	9471947	84.7	3.12	0.33
	Zape 05	18007010	87.6	1.3	0.22
	Zape 28	23920143	88.4	0.31	0.28

Table 2. Summary of reads

Reference Genome	Genus Present in QIIME/MIDAS	Species Present in QIIME/MIDAS	Genus Present in HMP Gut	Species Present in HMP Gut
Anaerostipes hadrus DSM 3319	X		X	X
Bacteroides uniformis CL03T00C23	X	X	X	X
Blautia wexlerae DSM 19850	X	X	X	
Brachyspira aalborgi	X	X		
Catenibacterium mitsuokai DSM 15897	X	X	X	X
Clostridium bartlettii DSM 16795	X	X	X	X
Collinsella aerofaciens ATCC 25986	X	X	X	X
Coprococcus sp. ART55/1	X		X	
Desulfovibrio piger ATCC 29098	X	X	X	X
Streptococcus anginosus C238	X	X	X	X
Eubacterium eligens ATCC 27750	X	X		
Faecalibacterium prausnitzii A2-165	X	X	X	X
Meiothermus chliarophilus DSM 9957	X			
Lachnospira multipara ATCC 19207	X	X		
Oscillibacter sp. 1-3 genomic scaffold	X		X	
Phascolarctobacterium sp. YIT 12067	X	X	X	
Prevotella copri DSM 18205	X	X	X	X
Eubacterium siraeum DSM 15702	X	X	X	X
Ruminococcus bromii L2-63	X	X	X	
Streptococcus thermophilus LMG 18311	X		X	
Treponema succinifaciens DSM 2489	X	X	X	

Table 3. Summary of reference genomes.

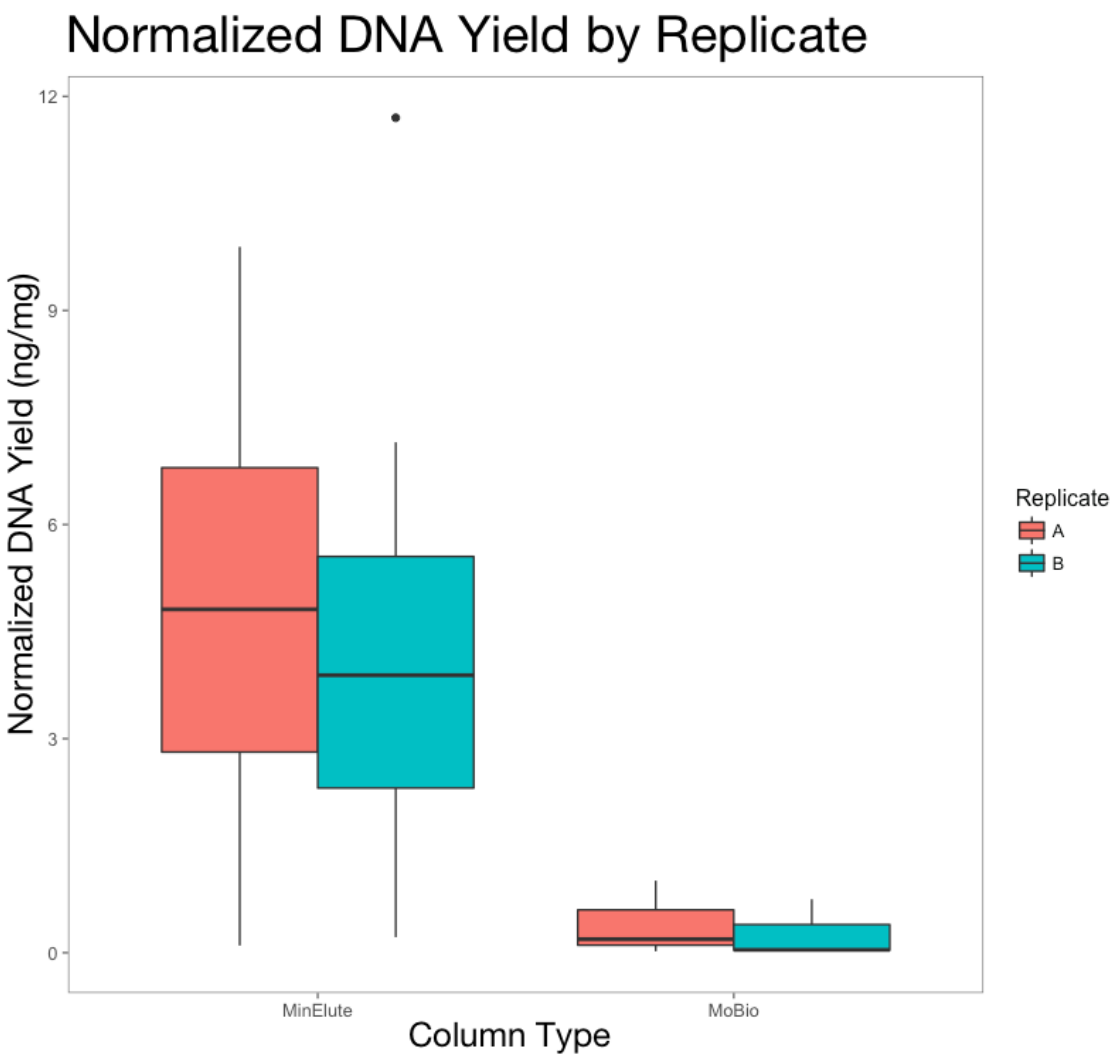

Figure 1. Normalized DNA yield by replicate. Both replicates 'A' and 'B' showed significant differences in the yield of normalized DNA by column type.

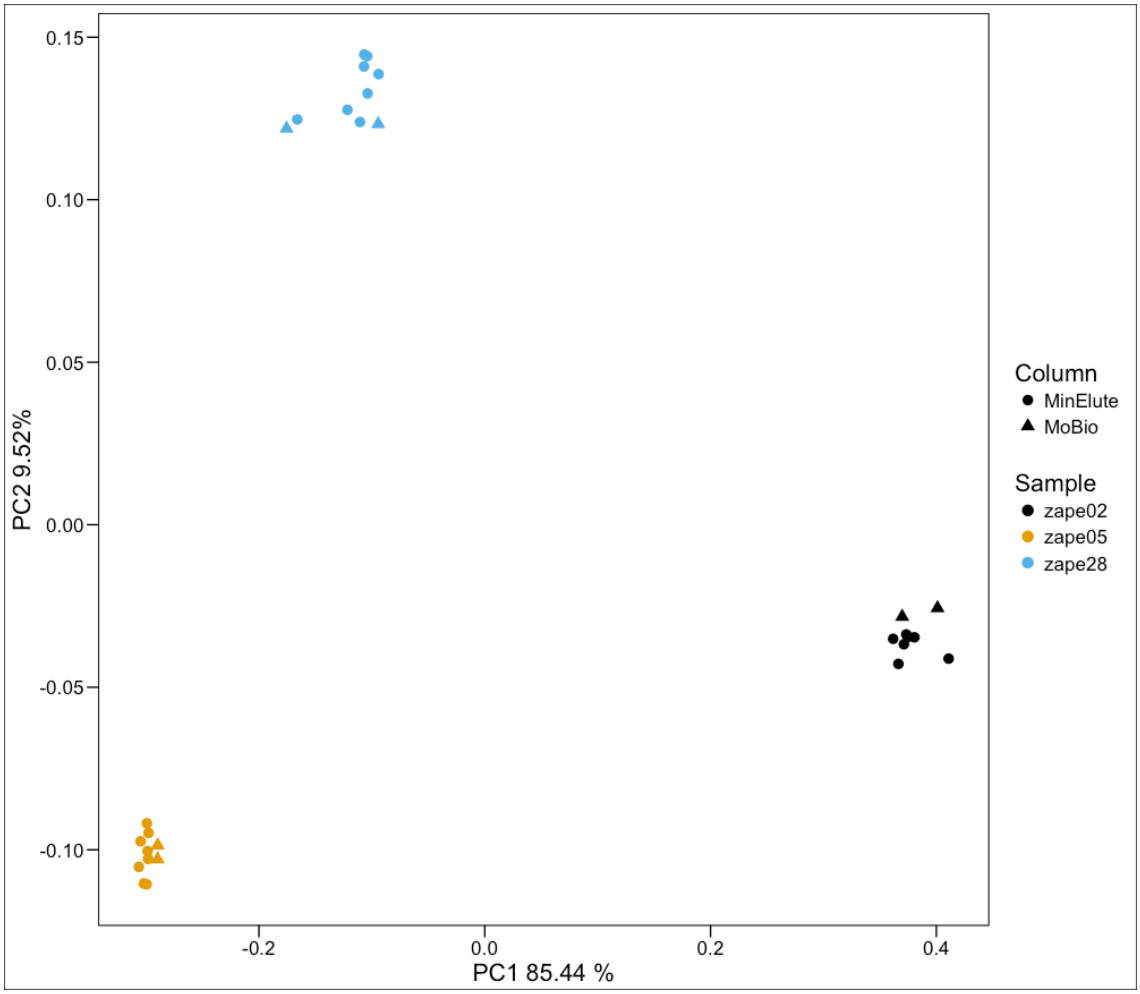

Figure 2. Weighted UniFraC β-diversity.

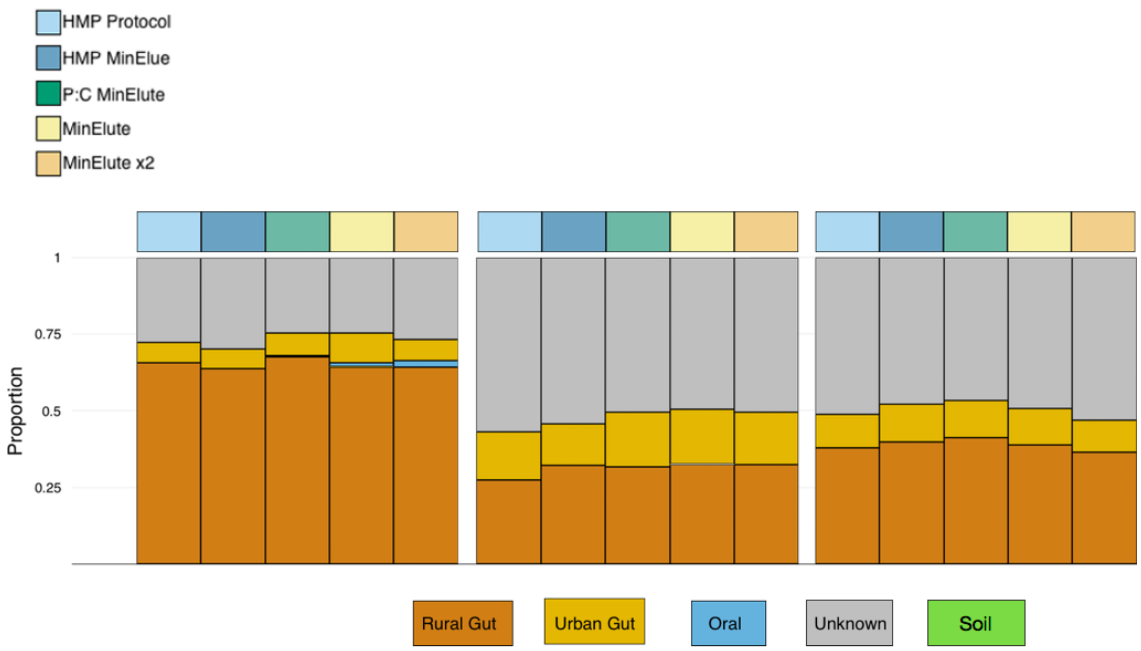

Figure 3. Microbial Community Source Tracking

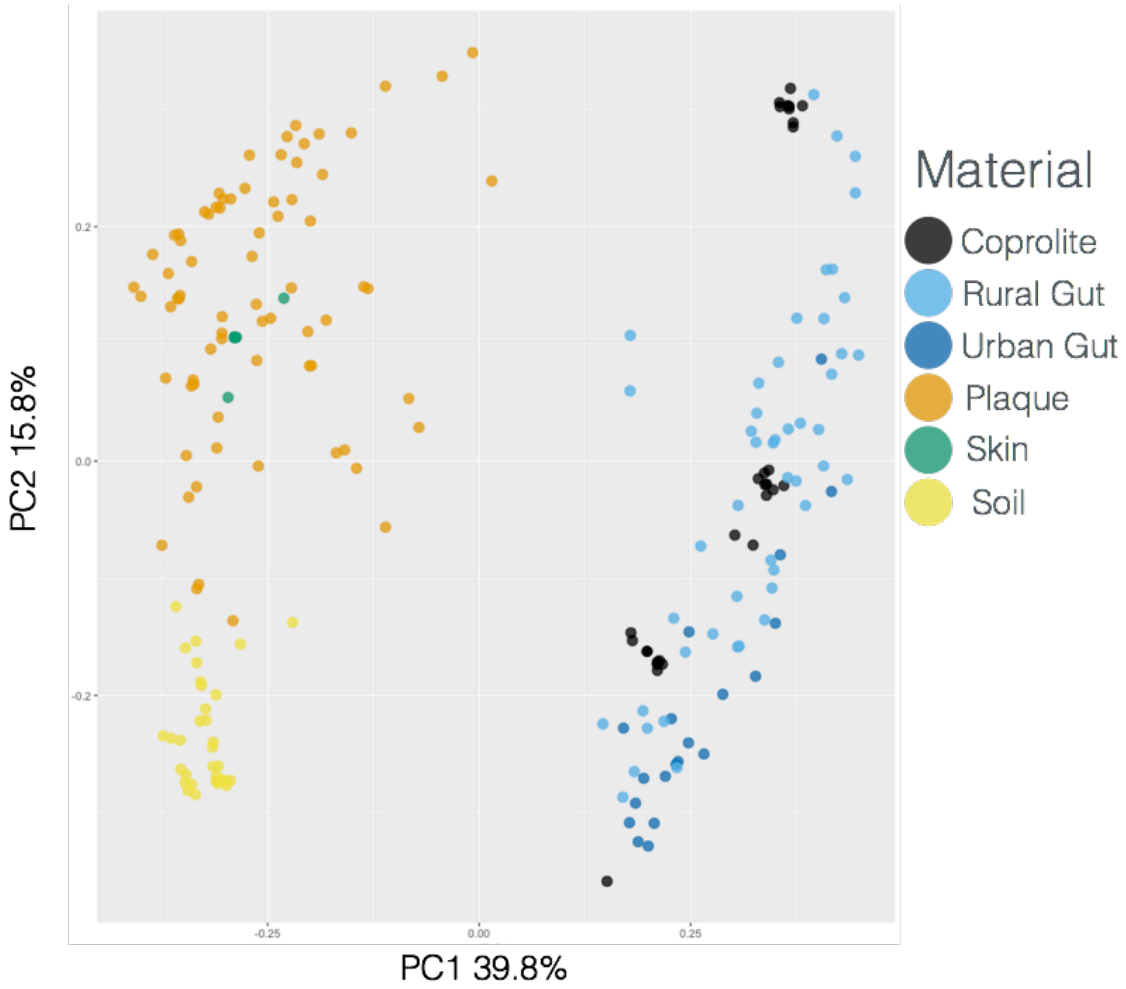

Figure 4. Weighted UniFrac on SourceTracker data

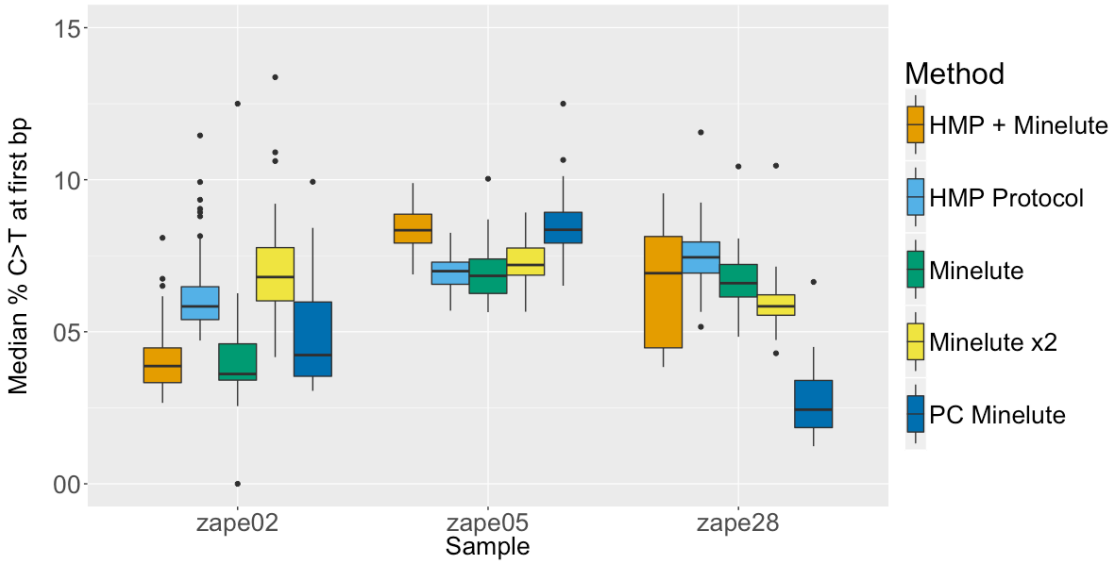

Figure 5. Distribution of DNA damage by method and sample

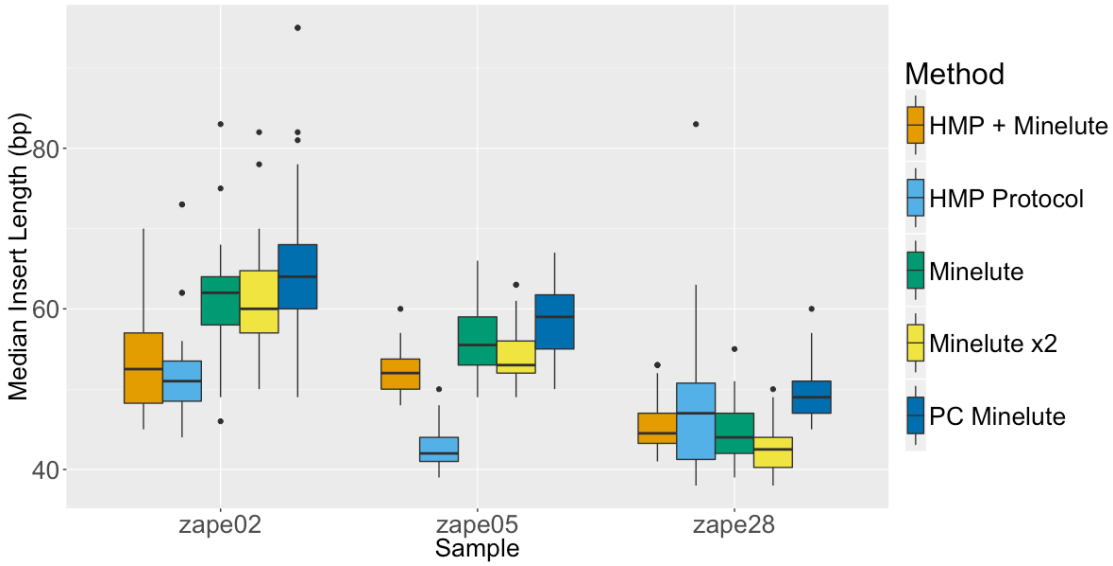

Figure 6. Distribution of insert lengths by sample and method

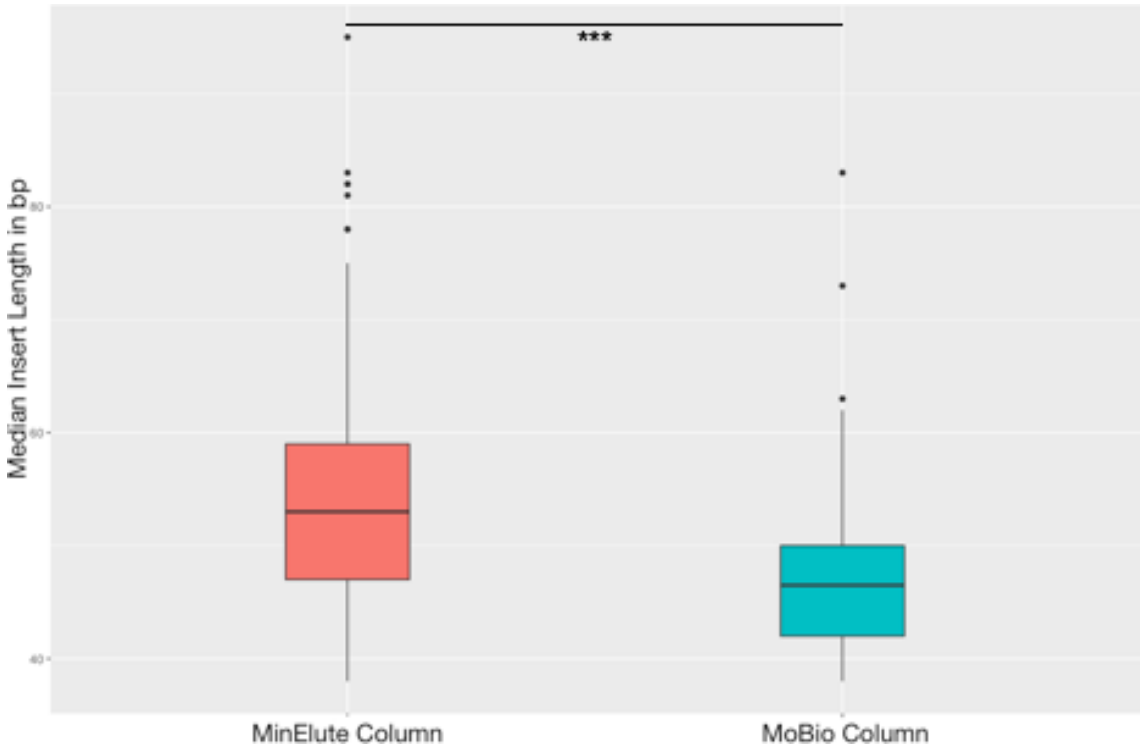

Figure 7. Comparison of insert lengths by column type

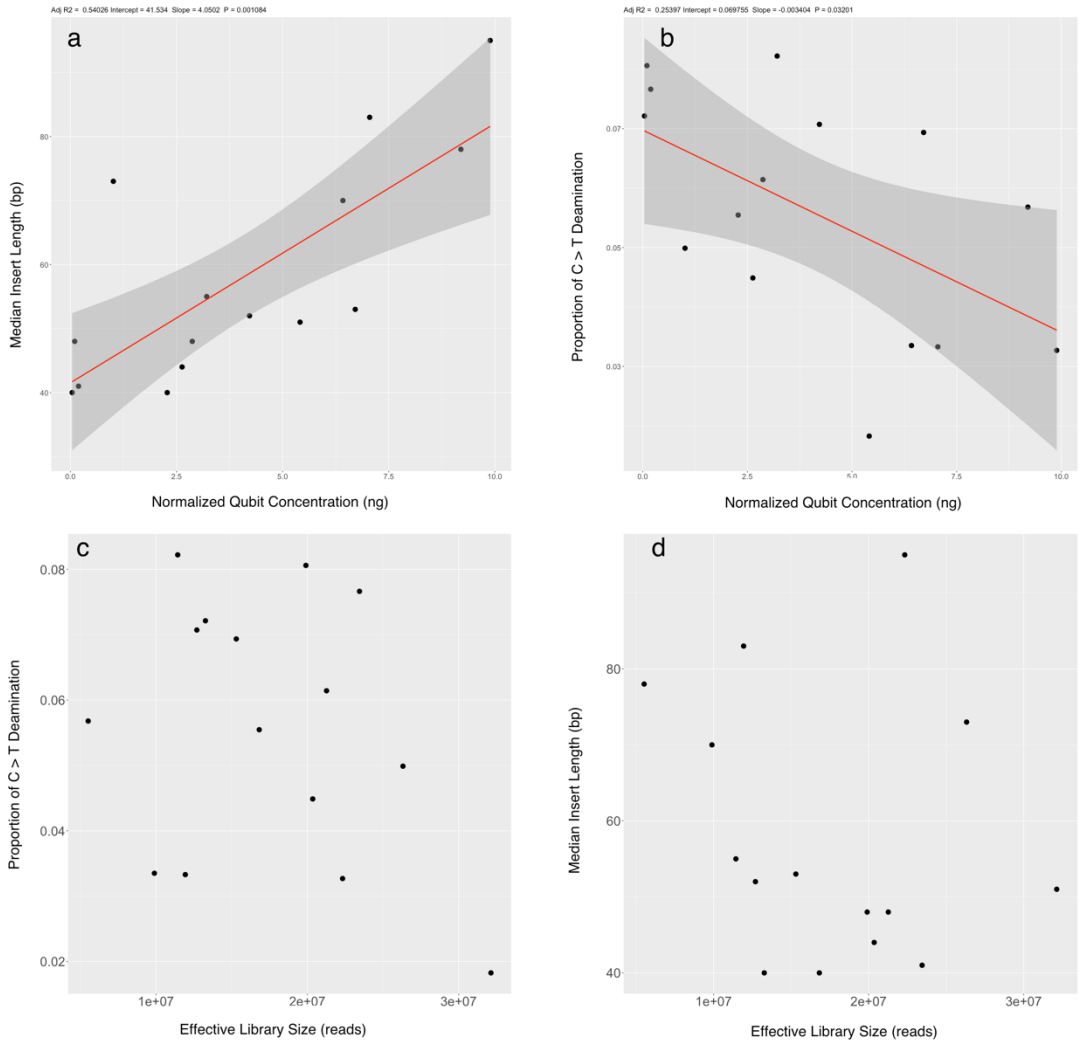

Figure 8. Linear modeling of experimental variables

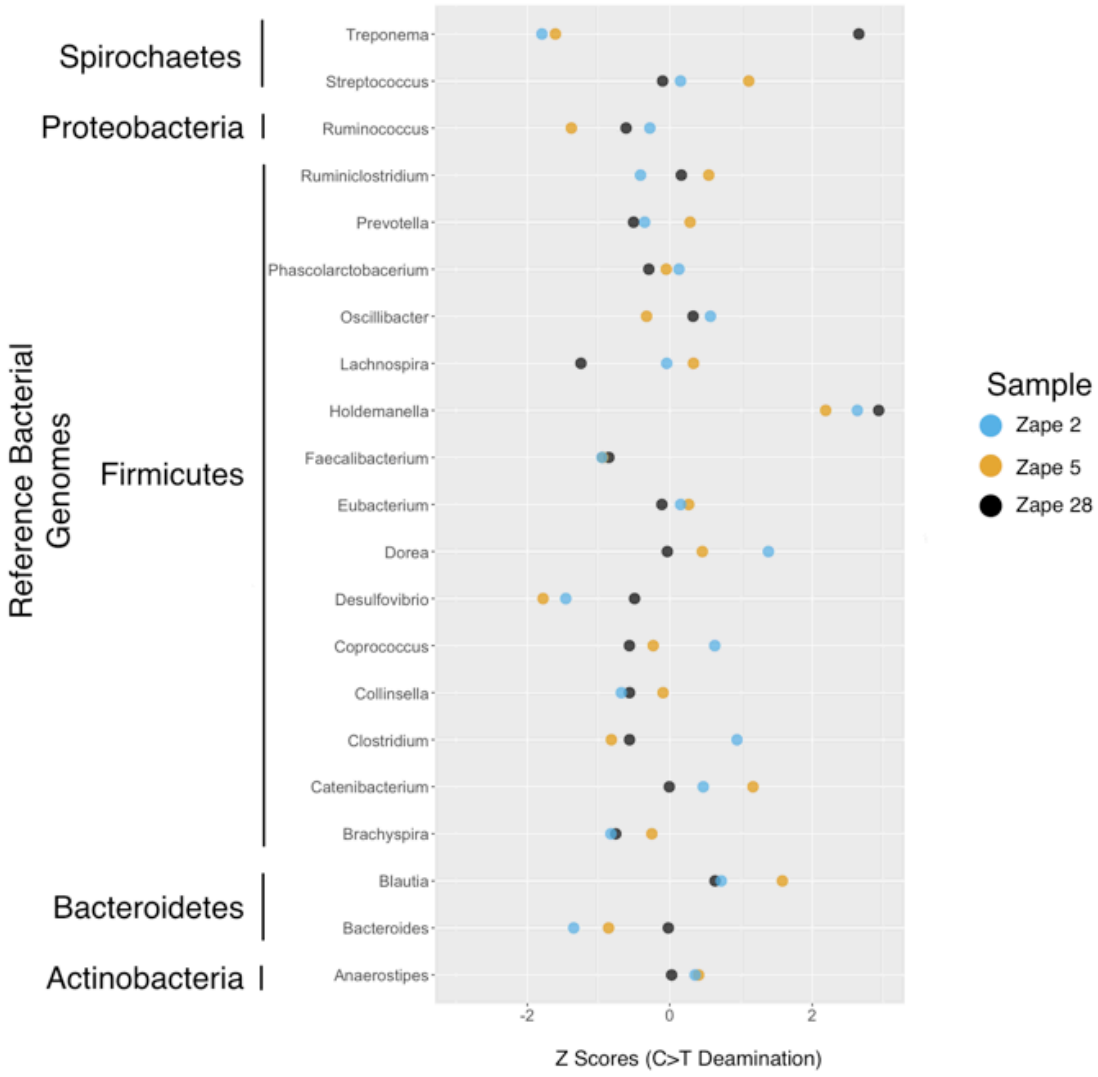

Figure 9. Variance of DNA damage by reference taxon

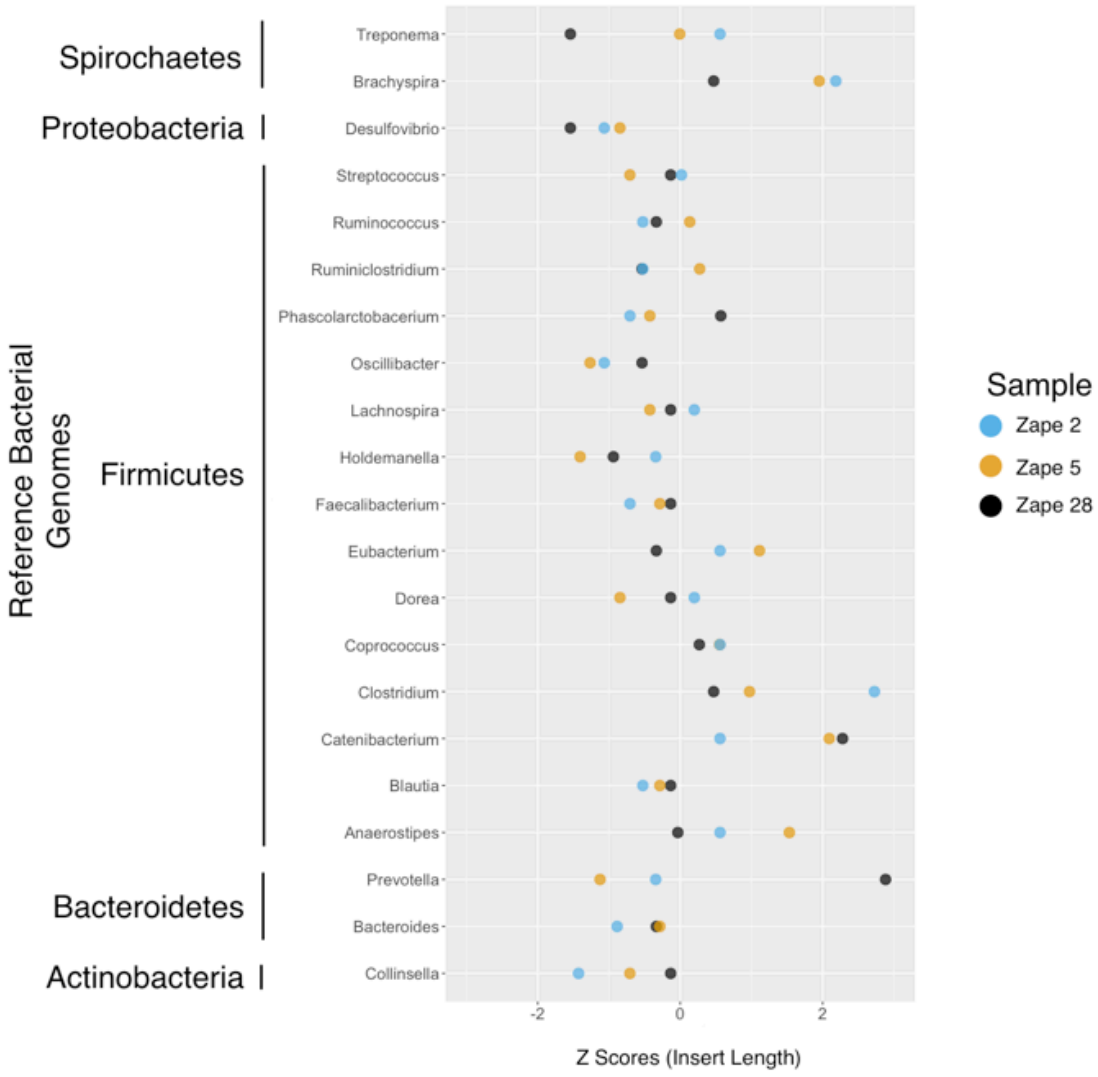

Figure 10. Variance of insert length by taxon

Chapter 4: Discussion

Microbial archaeology is a fast-growing field, as is its internal niche of ancient microbiome studies. Much of the work in reconstructing ancient microbiomes has focused on the human oral microbiome utilizing dental calculus as a rich matrix of ancient biomolecules, including DNA from commensal and pathogenic bacteria in the oral cavity (Warinner 2014, Metcalf 2014). Several studies have been published that utilize coprolites as a source for ancient DNA, but only a few of these have focused on reconstructing ancient microbial communities, opting instead to use the material as a source of endogenous DNA (Poinar et al. 2003, Bon et al. 2013, Wood et al. 2016). The studies which do examine the microbial communities from coprolites have employed a wide range of extraction and sequencing technologies, none of which have been verified to produce results that are compatible with the most widely-used methods for modern gut microbiome studies (Tito et al. 2012, Santiago-Rodriguez et al. 2015). As well-preserved human coprolites represent a rare and finite archaeological resource, it is important that the methods used to study them produce results that can be compared to modern data if we hope to understand the evolution of the human gut microbiome within the context of anthropological change.

In this study, I sought to compare five different methods for the extraction of ancient DNA from coprolites in order to determine their efficacy and accuracy for reconstructing the gut microbiomes of past peoples. Additionally, this experiment was designed to test whether data generated using methods specifically designed for aDNA are comparable with data generated through methods used for the recovery of DNA from modern human feces. The results indicate that methods specifically tailored to

aDNA are significantly better at recovering a higher quantity of DNA from coprolites. Though unsurprising given previous work showing that the recovery of aDNA requires modified techniques to overcome the challenges associated with it (Rohland and Hofreiter 2007, Dabney et al. 2013b), this finding recapitulates those studies with a novel material in coprolites, as opposed to bone or teeth.

The results of this study also support the notion that the methods geared towards recovery of aDNA do not alter the reconstructed microbial community when compared to the methods used for modern gut microbiome sampling. This finding is critical for the near future of ancient microbiome studies, as it provides the first evidence that data generated using protocols tailored for aDNA can be directly compared to data generated using methods for modern gut microbiome studies, such as those employed by the Human Microbiome Project. Without this compatibility between modern and ancient data, studies of the evolution of the human gut microbiome via coprolites as a proxy material would be impeded by the uncertainty present in the ancient data. The findings presented here represent the first dataset which can be said to verify that comparisons between modern and ancient gut microbiomes are possible. This is an important addition the existing body of research surrounding the human gut microbiome, an area of study which has historically focused on individuals from Western, industrialized societies, and been impeded by population ascertainment biases (Sankaranarayanan 2016). Recently, these studies have been expanded to include modern populations living more traditional lifestyles such as rural agriculturalists and hunter-gatherers (Obregon-Tito et al. 2015, Schnorr et al. 2014). Findings from these studies indicate that industrialization has caused significant taxonomic shifts in the human gut microbiome,

notably the exclusion of members of the genus *Treponema*. While these shifts are apparent in living populations, almost nothing is known about the ancestral state of the human gut microbiome (Warinner et al. 2015).

The findings of the present study provide clear evidence that obtaining knowledge about the diversity of the ancestral human microbiome is a real possibility, and lays the groundwork for standardization and best practices in future studies of the gut microbiome via microbial archaeology. Additionally, this study improves our understanding of the extraction of aDNA from coprolites in a way that has not been shown previously. Specifically, the data suggest that like other archaeological materials, the presence of aDNA within coprolites is governed largely by the preservation of the samples themselves. It is well-known that taphonomic conditions differentially impact the survival of aDNA within archaeological samples (Collins et al. 2002), and despite the mechanisms being poorly understood thus far, it is clear that certain samples will simply yield better results than others, even when they are recovered from the same site (Kistler et al. 2017). Our data support that this also holds true for coprolites, as each one used in this study was recovered from the same archaeological excavation. These findings present new questions about the taphonomic processes governing the survival of aDNA in coprolites, as unlike bone and teeth, feces does not represent a stark matrix contrast between the source material and its surroundings — typically soil. Future experiments to elucidate the processes by which DNA is preserved within feces long after it is deposited are warranted. Ideally, a longitudinal metagenomics study of feces deposited into different soil matrices should be performed. An additional approach to

resolving the ways in which DNA is preserved within feces is to employ the techniques used in this study on coprolites from unique archaeological environments.

The human gut microbiome represents arguably the most intricate and important symbiotic relationship that humans have, reflecting a deep co-evolutionary history (Muegge et al. 2011). Modern studies of the human gut microbiome continue to reveal novel ways in which this rich ecosystem is intertwined with human physiology and lifestyle, but are limited to providing a snapshot of evolution in action. Data representing past snapshots of the human gut microbiome are critical for understanding this environment from an evolutionary perspective, and the present study has shown that methods widely used for aDNA extraction in paleogenomics are able to generate data that allow for comparisons with modern studies.

References

Achtman M, Wagner M. 2008. Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol 6:431–440.

Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, Bradshaw CJ a, Townsend G, Sołtysiak A, Alt KW, Parkhill J, Cooper A. 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat Genet 45:450–455.

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway RN, Bunce M. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc Biol Sci 279:4724–33.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas A-S, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Casa P Della, Da ˛browski P, Duffy PR, Ebel A V, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. 2015. Population genomics of Bronze Age Eurasia. Nature 522:167–172.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E. 2005. Defining operational taxonomic units using DNA barcode data. Philos Trans R Soc Lond B Biol Sci 360:1935–1943.

Bon C, Berthonaud V, Maksud F, Labadie K, Poulain J, Artiguenave F, Wincker P, Aury JM, Elalouf JM. 2012. Coprolites as a source of information on the genome and diet of the cave hyena. Proc Biol Sci 279:2825–2830.

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest S a, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur a K, Guichon R a, Wolfe Steadman DL, Cook DC, Niemann S, Behr M a, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, 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.

Bos KI, Schuenemann VJ, Golding GB, Burbano H a, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJD, Herring DA, Bauer P, Poinar HN, Krause J. 2011. A draft genome of Yersinia pestis from victims of the Black Death. Nature 478:506–10.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A 104:14616–14621.

Briggs AW, 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.

Brosius J, Dull TJ, Sleeter DD, Noller HF. 1981. Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli. J Mol Biol 148:107–127.

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Res 35:5717–5728.

Cambies. 1949. Parodontose et cure Thermale. Odontol Rev 71:447–452.

Cano RJ, Borucki MK. 1995. Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. Science 268:1060–1064.

Cohen SN, Chang ACY, Boyert HW, Hellingt RB. 1973. Construction of Biologically Functional Bacterial Plasmids In Vitro (R factor/restriction enzyme/transformation/endonuclease/antibiotic resistance). Proc Natl Acad Sci 70:3240–3244.

Collins MJ, Nielsen–Marsh CM, Hiller J, Smith CI, Roberts JP, Prigodich R V, Wess TJ, Csapo J, Millard AR, Turner–Walker G. 2002. The survival of organic matter in bone: a review. Archaeometry 44:383–394.

Cooper A, Poinar HN. 2000. Ancient DNA. Nat Genet 2:353.

Cooper A, Poinar HN. 2000. Ancient DNA: do it right or not at all. Science 289:1139.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial community variation in human body habitats across space and time. Science (80-) 326:1694.

Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga J-L, Meyer M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A 110:15758–63.

Dabney J, Meyer M, Pääbo S. 2013. Ancient DNA damage. Cold Spring Harb Perspect Biol 5:a012567.

Danna K, Nathans D. 1971. Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. Proc Natl Acad Sci 68:2913–2917.

Dayhoff MO. 1979. Atlas of Protein Sequence and Structure. Volume 5,:1–414.

Dominguez-Bello MG, De Jesus-Laboy KM, Shen N, Cox LM, Amir A, Gonzalez A, Bokulich NA, Song SJ, Hoashi M, Rivera-Vinas JI, Mendez K, Knight R, Clemente JC. 2016. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nat Med 22:250–253.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A 107:11971–11975.

Faith DP. 1992. Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10.

Gaillard C, Strauss F. 1990. Ethanol precipitation of DNA with linear polyacrylamide as carrier. Nucleic Acids Res 18:378.

Gao Z, Tseng CH, Strober BE, Pei Z, Blaser MJ. 2008. Substantial alterations of the cutaneous bacterial biota in psoriatic lesions. PLoS One 3:e2719.

Gilichinsky DA, Wilson GS, Friedmann EI, McKay CP, Sletten RS, Rivkina EM, Vishnivetskaya TA, Erokhina LG, Ivanushkina NE, Kochkina GA, Shcherbakova VA, Soina VS, Spirina E V, Vorobyova EA, Fyodorov-Davydov DG, Hallet B, Ozerskaya SM, Sorokovikov VA, Laurinavichyus KS, Shatilovich A V, Chanton JP, Ostroumov VE, Tiedje JM. 2007. Microbial populations in Antarctic permafrost: biodiversity, state, age, and implication for astrobiology. Astrobiology 7:275–311.

Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011. mapDamage: Testing for damage patterns in ancient DNA sequences. Bioinformatics 27:2153–2155.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova L V, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A draft sequence of the Neandertal genome. Science 328:710–22.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA. 2009. Topographical and temporal diversity of the human skin microbiome. Science 324:1190–2.

Grice EA, Segre JA. 2011. The skin microbiome. Nat Rev Microbiol 9:244–53.

Hagelberg E, Hofreiter M, Keyser C. 2015. Introduction. Ancient DNA: the first three decades. Philos Trans R Soc Lond B Biol Sci 370:20130371.

Handt O, Höss M, Krings M, Pääbo S. 1994. Ancient DNA: Methodological challenges. Experientia 50:524–529.

Harper KN, Armelagos GJ. 2013. Genomics, the origins of agriculture, and our changing microbe-scape: time to revisit some old tales and tell some new ones. Am J Phys Anthropol 152:135–52.

Harper KN, Zuckerman MK, Harper ML, Kingston JD, Armelagos GJ. 2011. The origin and antiquity of syphilis revisited: An Appraisal of Old World pre-Columbian evidence for treponemal infection. Am J Phys Anthropol 146:99–133.

He X, Shi W. 2009. Oral Microbiology: Past, Present and Future. Int J Oral Sci 1:47–58.

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 312:282–284.

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S. 1996. DNA damage and DNA sequence retrival from ancient tissues. Nucleic Acids Res 24:1304–1307.

Höss M, Pääbo S. 1993. DNA extraction from pleistocene bones by a silica-based purification method. Nucleic Acids Res 21:3913–3914.

Huss J. 2014. Methodology and ontology in microbiome research. Biol Theory 9:392–400.

Huttenhower C, Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–14.

Iosif Lazaridis, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H. Sudmant, Joshua G. Schraiber, Sergi Castellano, Mark Lipson, Bonnie Berger, Christos Economou, Ruth Bollongino, Qiaomei Fu, Kirsten I. Bos, Susanne NR et al. 2014. Ancient human genomes suggest three ancestralpopulations for present-day European. Nature 513:1–20.

Jeffreys AJ. 1984. Raising the Dead and Buried. Nature 312:198.

Jenkins DL, Davis LG, Stafford Jr. TW, Campos PF, Hockett B, Jones GT, Cummings LS, Yost C, Connolly TJ, Yohe 2nd RM, Gibbons SC, Raghavan M, Rasmussen M, Paijmans JL, Hofreiter M, Kemp BM, Barta JL, Monroe C, Gilbert MT, Willerslev E. 2012. Supplementary Materials for Clovis age Western Stemmed projectile points and human coprolites at the Paisley Caves. Science (80-) 337:223–228.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29:1682–1684.

Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. 2014. Metatranscriptomics of the human oral microbiome during health and disease. MBio 5:e01012-14.

Journal A, Physical OF. 1998. Technical note : improved DNA extraction from ancient bones using silica-based spin columns . Am J Phys Anthropol Technical Note : Improved DNA Extraction From Ancient Bones. Am J Phys Anthropol 543:539–543.

Kistler L, Ware R, Smith O, Collins M, Allaby RG. 2017. A new model for ancient DNA decay based on paleogenomic meta-analysis. bioRxiv.

Krishnan K, Chen T, Paster B. 2016. A practical guide to the oral microbiome and its relation to health and disease. Oral Dis 23:276–286.

Lamont RF, Sobel JD, Akins RA, Hassan SS, Chaiworapongsa T, Kusanovic JP, Romeroa R. 2011. The vaginal microbiome: New information about genital tract flora using molecular based techniques. BJOG An Int J Obstet Gynaecol 118:533–549.

Lederberg BJ, McCray AT. 2001. ’ Ome Sweet ’ Omics-- A Genealogical Treasury of Words. Sci [Internet] 15:8. Available from: http://www.the-scientist.com/?articles.view/articleNo/13313/title/-Ome-Sweet--Omics---A-Genealogical-Treasury-of-Words/

Lederberg J. 2003. Of Men and Microbes. New Perspect Q 20:52–55.

Leewenhoek A. 1684. An abstract of a letter from Mr. Antonie van Leewenhoek, dated Sep.17, 1693. Containing some microscopical observations, about animals in the scurf of the teeth... Philos Trans 14:568–574.

Lewis CM, Obregón-Tito A, Tito RY, Foster MW, Spicer PG. 2012. The Human Microbiome Project: Lessons from human genomics. Trends Microbiol 20:1–4.

Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6:776–788.

Ley RE, Hamady M, Lozupone CA, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. Science 320:1647–51.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. Nature 362:709–715.

Lozupone C, Knight R. 2005. UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. Appl Environ Microbiol 71:8228–8235.

Luciani S, Fornaciari G, Rickards O, Labarga CM, Rollo F. 2006. Molecular characterization of a pre-Columbian mummy and in situ coprolite. Am J Phys Anthropol 129:620–629.

Ma B, Forney LJ, Ravel J. 2012. Vaginal microbiome: rethinking health and disease. Annu Rev Microbiol 66:371–89.

Marchesi JR, Ravel J. 2015. The vocabulary of microbiome research: a proposal. Microbiome 3:31.

McCoy CO, Matsen F a. 2013. Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth. PeerJ 1:e157.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey H V, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ. 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci 110:3229–3236.

Metcalf JL, Xu ZZ, Weiss S, Treuren W Van, Hyde ER, Song SJ, Amir A, Larsen P, Sangwan N, Haarmann D, Humphrey GC, Ackermann G, Thompson LR, Lauber C, Bibat A, Nicholas C, Gebert MJ, Petrosino JF, Reed SC, Lynne AM, Bucheli SR, Carter DO. 2016. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science 351:158–62.

Metcalf JL, Wegener Parfrey L, Gonzalez A, Lauber CL, Knights D, Ackermann G, Humphrey GC, Gebert MJ, Van Treuren W, Berg-Lyons D, Keepers K, Guo Y, Bullard J, Fierer N, Carter DO, Knight R. 2013. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. Elife 2:e01104.

Moeller AH, Li Y, Mpoudi Ngole E, Ahuka-Mundeke S, Lonsdorf E V, Pusey AE, Peeters M, Hahn BH, Ochman H. 2014. Rapid changes in the gut microbiome during human evolution. Proc Natl Acad Sci U S A 111:16431–5.

Mohr JL. 1952. Protozoa as Indicators of Pollution. Sci Mon 74:7–9.

Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science (80-) 332:970–974.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51:263–273.

O'Hara AM, Shanahan F. 2006. The gut flora as a forgotten organ. EMBO Rep 7:688–693.

Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, Zech Xu Z, Van Treuren W, Knight R, Gaffney PM, Spicer P, Lawson P, Marin-Reyes L, Trujillo-Villarroel O, Foster M, Guija-Poma E, Troncoso-Corzo L, Warinner C, Ozga AT, Lewis CM. 2015. Subsistence strategies in traditional societies distinguish gut microbiomes. Nat Commun 6:6505.

Ochman H, Worobey M, Kuo CH, Ndjango JBN, Peeters M, Hahn BH, Hugenholtz P. 2010. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol 8:e1000546.

Ozga AT, Colon MN, Honap T, Sankaranarayanan K, Hofman C, Milner G, Lewis Jr C, Stone A, Warinner C. 2016. Ancient dental calculus as a reservoir of whole human mitogenomes. In: American Journal of Physical Anthropology. Vol. 159. WILEY-BLACKWELL 111 RIVER ST, HOBOKEN 07030-5774, NJ USA. p 245–246.

Pääbo S. 1985. Molecular-Cloning of Ancient Egyptian Mummy DNA. Nature 314:644–645.

Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, Sahasrabudhe A, Dewhirst FE. 2001. Bacterial diversity in human subgingival plaque. J Bacteriol 183:3770–3783.

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, 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, Roger Little A, Peavy H, Pontzer C, Portnoy M, Sayre MH,

Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M. 2009. The NIH Human Microbiome Project. Genome Res 19:2317–2323.

Peterson SN, Snesrud E, Liu J, Ong AC, Kilian M, Schork NJ, Bretz W. 2013. The Dental Plaque Microbiome in Health and Disease. PLoS One 8:e58487.

Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB, Hallin S. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol 8:523–529.

Poinar H, Kuch M, McDonald G, Martin P, Pääbo S. 2003. Nuclear gene sequences from a late Pleistocene sloth coprolite. Curr Biol 13:1150–1152.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, Green JL, Green LE, Killham K, Lennon JJ, Osborn AM, Solan M, van der Gast CJ, Young JPW. 2007. The role of ecological theory in microbial ecology. Nat Rev Microbiol 5:384–392.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci 108:4680–4687.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov M V, Derevianko AP, Hublin J-J, Kelso J, Slatkin M, Pääbo S. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053–1060.

Reinhard KJ. 1988. Recovery of parasite remains from coprolites and latrines: aspects of paleoparasitological technique Paleoparasitologia no Brasil. :158–108.

Reinhard KJ, Vaughn B. 1992. Coprolite Analysis: A Biological Perspective on Archaeology. Archaeol Method Theory 4:245–288.

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. 2015. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos Trans R Soc Lond B Biol Sci 370:20130624.

Rohland N, Hofreiter M. 2007. Ancient DNA extraction from bones and teeth. Nat Protoc 2:1756–1762.

Rosan B, Lamont RJ. 2000. Dental plaque formation. Microbes Infect 2:1599–1607.

Salo WL, Aufderheide AC, Buikstrat J, Holcomb TA. 1994. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. Microbiology 91:2091–2094.

Sanger F, Nicklen S, Coulson a R. 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–7.

Sankaranarayanan K, Lowenstein TK, Timofeeff MN, Schubert B a, Lum JK. 2014. Characterization of ancient DNA supports long-term survival of Haloarchaea. Astrobiology 14:553–60.

Sankaranarayanan K, Ozga AT, Warinner C, Tito RY, Obregon-Tito AJ, Xu J, Gaffney PM, Jervis LL, Cox D, Stephens L. 2015. Gut microbiome diversity among Cheyenne and Arapaho individuals from western Oklahoma. Curr Biol 25:3161–3169.

Santiago-Rodriguez TM, Fornaciari G, Luciani S, Dowd SE, Toranzos GA, Marota I, Cano RJ. 2015. Gut microbiome of an 11 th century AD Pre-Columbian Andean mummy. PLoS One 10:e0138135.

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.

Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turroni S, Biagi E, Peano C, Severgnini M, Fiori J, Gotti R, De Bellis G, Luiselli D, Brigidi P, Mabulla A, Marlowe F, Henry AG, Crittenden AN. 2014. Gut microbiome of the Hadza hunter-gatherers. Nat Commun 5:3654.

Schuenemann VJ, Bos K, DeWitte S, Schmedes S, Jamieson J, Mittnik A, Forrest S, Coombes BK, Wood JW, Earn DJD, White W, Krause J, Poinar HN. 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. Proc Natl Acad Sci U S A 108:E746-52.

Sekirov I, Russell S, Antunes L. 2010. Gut microbiota in health and disease. Physiol Rev 90:859–904.

Tito RY, Macmil S, Wiley G, Najar F, Cleeland L, Qu C, Wang P, Romagne F, Leonard S, Ruiz AJ, Reinhard K, Roe BA, Lewis Jr. CM. 2008. Phylotyping and functional analysis of two ancient human microbiomes. PLoS One 3:e3703.

Tito RY, Knights D, Metcalf J, Obregon-Tito AJ, Cleeland L, Najar F, Roe B, Reinhard K, Sobolik K, Belknap S, Foster M, Spicer P, Knight R, Lewis CM. 2012. Insights from Characterizing Extinct Human Gut Microbiomes. PLoS One 7:e51146.

Wade WG. 2013. The oral microbiome in health and disease. Pharmacol Res 69:137–143.

Warinner C, Rodrigues JFM, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-García DC, Eppler E, Seiler R, Hansen LH, Castruita JAS, Barkow-Oesterreicher S, Teoh KY, Kelstrup CD, Olsen J V, Nanni P, Kawai T, Willerslev E, von Mering C, Lewis CM, Collins MJ, Gilbert MTP, Rühli F, Cappellini E. 2014. Pathogens and host immunity in the ancient human oral cavity. Nat Genet 46:336–44.

Warinner C, Speller C, Collins MJ. 2015. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. Philos Trans R Soc Lond B Biol Sci 370:20130376.

Warinner C, Speller C, Collins MJ, Lewis CM. 2015. Ancient human microbiomes. J Hum Evol 79:125–136.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 6S Ribosomal DNA Amplification for Phylogenetic Study. J Bacteriol 173:697–703.

Weyrich LS, Dobney K, Cooper A. 2015. Ancient DNA analysis of dental calculus. J Hum Evol 79:119–124.

Wilbur AK, Bouwman AS, Stone AC, Roberts CA, Pfister LA, Buikstra JE, Brown TA. 2009. Deficiencies and challenges in the study of ancient tuberculosis DNA. J Archaeol Sci 36:1990–1997.

Willerslev E, Hansen AJ, Rønn R, Brand T, Barnes I, Wiuf C, Gilichinsky D, Mitchell D, Cooper A. 2010. Long-term persistence of bacterial DNA. Curr Biol 14:R9–R10.

Witkin SS, Linhares IM, Giraldo P. 2007. Bacterial flora of the female genital tract: function and immune regulation. Best Pract Res Clin Obstet Gynaecol 21:347–354.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74:5088–5090.

Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, Pechman K, Lewis BJ, Stahl D. 1975. Conservation of primary structure in 16S ribosomal RNA. Nature 254:83–86.

Wood JR, Crown A, Cole TL, Wilmshurst JM. 2016. Microscopic and ancient DNA profiling of Polynesian dog (kurī) coprolites from northern New Zealand. J Archaeol Sci Reports 6:496–505.

Ziesemer KA, Mann AE, Sankaranarayanan K, Schroeder H, Ozga AT, Brandt BW, Zaura E, Waters-Rist A, Hoogland M, Salazar-García DC, Aldenderfer M, Speller C, Hendy J, Weston DA, MacDonald SJ, Thomas GH, Collins MJ, Lewis CM, Hofman C, Warinner C. 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. Sci Rep 5:16498.

Zink A, Haas CJ, Reischl U, Szeimies U, Nerlich AG. 2001. Molecular analysis of skeletal tuberculosis in an ancient egyptian population. J Med Microbiol 50:355–366.