PALEOPATHOLOGICAL AND MICROBIOLOGICAL INVESTIGATIONS OF DENTAL HEALTH IN AMERICA SINCE 1890

By

LASHANDA RENA WILLIAMS

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Ecology and Evolution

Written under the direction of

Siobain Duffy

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2019

ABSTRACT OF THE DISSERTATION

PALEOPATHOLOGICAL AND MICROBIOLOGICAL INVESTIGATIONS OF DENTAL HEALTH IN AMERICA SINCE 1890

By LASHANDA RENA WILLIAMS

Dissertation Director:

Siobain Duffy

Human evolutionary history spans millions of years and is marked by several periods of cultural innovation. The most recent of which is the industrial revolution, which has impacted human health and biology. While modern sociocultural change has been well studied for its effects on systemic diseases, its effects on dental health in America has not well understood. Here, we use a multidisciplinary approach to examine the effects of modern sociocultural change on oral health and diseases. First, we combined dental paleopathology and modern dental methods to examine the dental health of late 19th century and early 20th century Americans and found that the dental health of studied American populations was poorer than people living today. Also, while there are statistical significant differences in root exposure, calculus abundance, and attrition, the overall dental health of this diverse dataset was the similar to their contemporaries, despite differences in socioeconomic status.

Secondly, we conducted a comprehensive metagenomics study on the Americans who died between 1895-1950 and we observed clear distinctions between modern dental plaque communities and historical calculus microbiota.

One prominent member of the historical calculus metagenome is *Methanobrevibacter oralis*, dominant in 41/43 of the samples analyzed. Like in the previous chapter, here I observed no distinct clustering patterns of microbiota and *M. oralis* pan-gene family content by age, “race”, and geography. The role of *M. oralis* in the modern oral cavity is that of a “co-pathogen” associated with periodontitis, however further analysis using historical reconstructions and modern genomic data will shed light on the pathogenic potential of this archaeon over time.

Lastly, as a control for all dental calculus microbiome studies, I compared the microbial communities of dental plaque to that of dental calculus and found that dental calculus is its own distinct microbial community and preserves a similar alpha diversity to other oral microhabitats. This pattern was observed when comparing a subject’s dental calculus to their own dental plaque and when comparing a subject’s dental calculus to HMP nasal, oral, and throat microbial communities. At the genus level, dental calculus harbors bacteria similar to that found in plaque, the differences reside at the species level. We observed 2 unique differentially abundant taxa within dental calculus, *Neisseria elongata* and *Cardiobacterium hominis*. Due to differentially abundant taxa and less preserved human DNA, ongoing ancient microbiome research has a bias towards recovering certain taxa and less human DNA will be available to study over time.

Acknowledgements

First, I would like to thank my family and friends who I love deeply and who have kept me grounded throughout this journey. Ma, Lil' B, Grandma La, Godmother Terri and the rest of my god-family, I don't know who nor where I'd be without you all. I'd also like to thank my closest friends Mike, Aja, Diana, Lele, and Natasia for being amazing moral and emotion support, for listening to my rants about graduate school, and consistently encouraging me along the way. I am so grateful to have a fantastic group of colleagues and dear friends from Duffy Lab 1.0 through 3.0 who've provided support, feedback, and encouragement on everything from fellowship essays to oral presentations, and for enjoyable discussions on everything from science to music to food.

I would have not considered graduate school or a career in science without the mentorship of two of my professors during my undergraduate, Dr. Christopher Brooks and Dr. Amiee Potter. Thank you Dr. Brooks for giving me the opportunity to teach and do research within the field of anthropology and igniting the desire to pursue a career in academia. Thank you Dr. Potter for being an inspiration to me, encouraged me to pursue biological anthropology, and mentored me over the years.

I'd like to thank my primary advisor, Dr. Siobain Duffy, for seeing my passion to pursue evolutionary medicine, and allowing me to pursue my interests no matter how far they veered from yours. I'd like to thank Dr. Carla Cugini who's support, enthusiasm, and excitement about my research has been infectious, invigorating, and hasn't waned since day 1. I'd also like to thank the remaining members of my dissertation committee, Dr. Rob Scott, Dr. Courtney Hofman, and Dr. Erin Vogel for their feedback and guidance during the conception and development of my dissertation research. I'd also like to thank

Dr. Sabrina Sholts, my advisor during my Smithsonian fellowship, for your continued mentorship and your support. Additionally, I'd like to thank Dr. Ken Markowitz for your encouragement and guidance throughout this process.

During my time at Rutgers, I have had the opportunity to mentored two phenomenal undergraduate students, Jessica Grandjean and Sherina Lawrence. I'd like to thank Jessica for supporting me during sampling trips and for exploring the oral microbiome research along with me. I'd also like to thank Sherina for delving into historical oral pathogens within my data and parsing through what seemed like dozens of spreadsheets to lay the foundation for one of my dissertation chapters.

I'd also like to thank all of the administrative support at ENR, especially Marsha Morin and Jennifer Scheck, for being available, congenial, and incredibly helpful to me over the last 5 years.

Gradfund has also been a great support and their mentorship has taught me how to successfully write and complete grant and fellowship applications. My mentors at Gradfund were the first to hear my dissertation research ideas and provided valuable feedback in the very beginning stages. Without your mentorship, I wouldn't have been able to fulfill my vision for my dissertation research.

Last, but not least, I'd like to thank the organizations that have funded my dissertation research: the National Science Foundation, the Smithsonian Institution Predoctoral Fellowship program, Rutgers' School of Graduate Studies, the Center for Human Evolutionary Studies, and the Department of Ecology and Evolution.

TABLE OF CONTENTS

Please see attached spreadsheet for the supplementary data files marked by an asterisk(*)

CHAPTER 1: INTRODUCTION

Human biological and cultural evolution spans millions of years, and has been significantly impacted by technological advancements. Some of the major cultural innovations include the invention of stone tools, agriculture, and most recently, the industrial revolution. These cultural changes have impacted human health and biology. Stone tools, for instance, are linked to increases in cognitive capacity [1]. Prior to the development of agriculture, the age of pre-Neolithic cultures is referred to as the Paleolithic Baseline [2]. During the Paleolithic Baseline, humans lived in small hunter-gatherer communities, ate a varied diet, and experienced low mortality and fertility rates [2]. The rise of agriculture and the attendant higher populations that could be supported by farming is associated with the first epidemiological transition (Age of Pestilence and Famine)[3]. The industrial revolution ushered in the Age of Receding Pandemics, as improved public sanitation and the development of effective medical interventions such as vaccines decreased childhood mortality and increased life expectancy [2, 3]. These health advances occurred despite industrialized food production leading to processed and packaged food with altered nutritional value [4].

The end of the industrial revolution, at the end of World War II, marks the beginning of the third epidemiological transition, the Age of Degenerative and Man-Made Diseases. It is so named because the causes of death shifted from infectious diseases to chronic, degenerative diseases [5]. For instance, in 1900, the leading causes of death in the United States were pneumonia, tuberculosis, diarrhea, enteritis, and ulceration of the intestines [6]. By 1945, the leading causes of death were diseases of the heart, cancer and other malignant tumors, and intracranial lesions of a vascular origin [6]. This shift was due to

improvements in diet and advancements in medicine, such as the discovery of antibiotics. The additional increases in life expectancy and declines in mortality also led to decreases in fertility [5, 7].

While the historical trajectory of systemic diseases has been well studied, the trajectory of oral and dental diseases is not well understood. The prevalence of dental diseases amongst Americans was not counted until 1959 -- after the third epidemiological transition had begun -- in the National Health Examination Survey [8]. This survey used the decay-missing-filled tooth method where the number of each tooth effected by each pathology are counted for every tooth present [8]. The prevalence of other dental conditions, such as periodontitis, oral debris, and dental calculus, were largely uncharacterized until recently in America the NHANES I (National Health and Nutrition Examination Survey) taken between 1950-1962 [9].

Unlike other microbiomes, the oral microbiome is directly associated with diseases that affect a majority of people in western and non-western societies alike [10]. Dental caries and periodontitis are two of the most common of all noncommunicable diseases in the United States [10]. The ecological plaque hypothesis is the most widely accepted hypothesis for how both caries and periodontitis occurs. The production of acid, which lowers the pH, leads to an ecological shift in resident microflora is the cause of demineralization associated with dental caries [11]. Plaque accumulation leads to increased inflammation and GCF (gingival cervical fluid) flow causes an ecological shift in resident microflora causing periodontal disease [11]. In good periodontal health, the microbiota of subgingival plaque is comprised of gram-positive bacteria that are predominantly facultative anaerobes [11]. In poor periodontal health, gram-negative bacteria that are predominantly anaerobes reside in subgingival

plaque. Additionally, individuals with periodontitis have greater diversity of microbial taxa in both subgingival and supragingival plaque [12].

Alternative hypotheses have been developed to explain the development of dental disease in the oral cavity. The keystone pathogen hypothesis suggests pathogens, such as *Porphyromonas gingivalis*, which is either present in low abundance or completely absent in health, causes the dysbiosis in the oral cavity by inducing inflammation [13]. Alternatively, oral microbiota are increasingly being viewed as pathobionts, commensals that can cause disease in certain environmental conditions [14]. These conditions include: immunodeficiency, antibiotic treatment, and the accumulation and/or infection of keystone pathogens,

Another attribute that distinguishes the oral microbiomes from other body microhabitats, is the ability for oral microflora to leave the oral cavity and take residency in other parts of the body [15]. Oral microbiota has been associated with over 100 systemic diseases localized in the nervous system, the cardiovascular system, the respiratory system, the endocrine system, the musclo-skeletal system, the gastrointestinal system, and the reproductive system [16, 17]. The mouth-body connection is an ancient concept, but advances in genomics and next-generation sequencing has led to a boom in research on the structure and function of the human microbiome, and how oral microbiota specifically can affect systemic health.

In this dissertation, I will investigate how the modern sociocultural and technological change of the third epidemiological transition has affected oral health, from the macroscopic to the microscopic level. Research comparing the oral microbiomes of urban western and rural non-western peoples suggests the modern diet, lifestyles, and behaviors have led to a change in oral microbial

ecology [18-20]. Oral bacteria of Amerindians, while similar at the phylum level, had lower diversity of genera and the increased carriage of uncharacterized bacterial taxa compared to other non-Amerindians [19]. This contrasts with a comparison of the oral microbiome of the Yanomami peoples of the Amazon to modern day Americans. The Yanomami and American subjects under study had similar levels of alpha diversity of bacterial phyla, however the Yanomami oral microbiome was distinct in beta diversity analysis and comprised of more oral taxa [20].

Ancient microbiome research is in its infancy, but has laid the foundations for investigating change in oral microbiota over time. This research has used dental calculus (tartar) to reconstruct the oral microbiomes of ancient humans and non-human mammals [21].

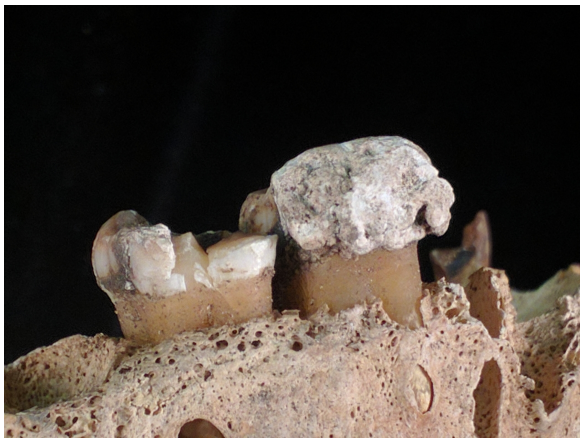


Figure 1.1: Dental calculus localized on the buccal surface of this individual's upper left first molar (HOW248337)

In a time-series of European populations from the Mesolithic (7,550-5,450 BP) to the modern era, Adler et al. (2013) sought to evaluate changes in human oral microbiota in response to dietary shifts between the Neolithic (7,400-4,000 BP) and the Industrial revolution (~1850). Using 16S sequencing, they characterized and compared the dominant taxa of each era [22] . This study had three major findings. First, the taxonomic composition of oral microbiota over time was similar across analyzed groups [22]. Second, despite these similarities, the oral microbiomes of modern Europeans have less biodiversity, relative to earlier time periods [22]. Lastly, modern Europeans have fewer bacteria associated with good health and an abundance of bacteria associated with dental caries and periodontal disease such as *Streptococcus mutans* and *Porphyromonas gingivalis* [22].

16S-based ancient microbiome research has continued to shed insight in the oral microbial populations of ancient peoples. In a time-series of Italians who lived between the Middle Ages through the Industrial Revolution, sociodemographic factors, such as social status and gender, were incorporated in analysis [23]. Operational taxonomic units (OTUs) over time were consistent and differences lay in their differential abundances [23]. Despite the taxonomic similarities, modern and oral microbial communities clustered away from each other in beta diversity analysis [23]. In a study of the oral microbiota of Saladoids, horticulturists who resided in Puerto Rico during the first century B.C., to modern day peoples uncovered differences in the composition of dental bacteria at the phylum level [24]. Additionally, the predicted functionome of ancient and modern dental calculus was also distinct, with cell motility, signal transduction, and biosynthesis of other secondary metabolites being unique features of the ancient dental calculus [24].

Whole genome shotgun sequencing (WGS) was first applied to ancient oral microbiome datasets by Warinner *et al.* (2014). 16S rRNA had been the gold standard in microbiome research due to its ubiquity across prokaryotes and its conserved and variable elements that also make it phylogenetically informative. However, it is less precise at low taxonomic levels, is not robust to sequencing errors, and is subject to amplification bias [25, 26]. Whole genome shotgun sequencing, in contrast, provides more comprehensive data about all organisms present within dental calculus, which is includes human, microbial, and diet-associated DNA [27]. This method was used to characterize the oral microorganisms and putative pathogens of four Europeans with mild to severe periodontal disease from a medieval monastic site in Germany (950-1200 CE) [27]. In addition to the bacteria, viral, and archaeal residents of the oral cavity, forty pathogens were described; these include *P. gingivalis*, and *T. denticola*[27]. Virulence factors, antibiotic resistance genes, and the reconstruction of periodontal pathogen *Tannerella forsythia* made also possible by WGS [27].

Farrer et al. (2018) combined archaeological, paleopathological, and metagenomics data in a study of dental calculus microbiota of Londoners pre- and post-Middle Ages. They found that oral geography (tooth type and surface) affects the microbiota and diversity of microflora recovered [28]. Oral diseases, such as abscesses, were also linked to oral geography, with their molar dataset being linked to microbiota variation [28]. The authors continued to explore ancient pathogens linked to oral diseases. They found members of *Prevotella* and *Streptococcus* were increased and *Porphyromonas gingivalis* and *Treponema denticola* were lower in Medieval Londoners with dental abscesses [28]. The proportion of Archaea, including *Methanobrevibacter* sp., were also

increased individuals with abscesses [28]. These findings suggest the etiology of oral diseases is multi-microbial in origin.

This dissertation will use multidisciplinary methods to investigate the evolution and microbial ecology of dental health and disease in American since 1890. In my second chapter, I will use traditional anthropological and modern dentistry methods to evaluate and calculate the prevalence of oral diseases in individuals who died at the turn of the century. In my third chapter, I will use WGS to characterize the oral metagenomes of geographically and racially diverse Americans using dental calculus. In my fourth chapter, I will conduct a taxonomic analysis of modern day dental calculus and evaluate its utility in reconstructing other oral microbial environments.

CHAPTER 2: DENTAL HEALTH AT THE CUSP OF THE THIRD EPIDEMIOLOGICAL TRANSITION

Abstract

Thorough investigation of the historical trajectory of a disease provides a better understanding of contemporary disease patterns. Our goal is to consider how the epidemiological transition model might be used to understand the oral health of past populations from the United States. This has the potential to provide necessary historical data, as the prevalence of oral disease in US residents was not measured systemically until the 1950s and 60s.

In this study, we evaluated dentition from museum collections to assess craniodental health of late 19th and early 20th century Americans using published scoring systems. Scores for each dental pathology were analyzed by Krustal-Wallis, ANOVA, and correlation analysis. Dental sockets were analyzed by collection, “race”, tooth number, and severity of dental pathologies.

We examined 1123 dental sockets from 46 geographically and racially diverse Americans from four skeletal collections. We observed a total of 157 caries, 37 of which were gross-gross occlusal caries. We also observed 182 cases of antemortem tooth loss and 417 teeth with ≥3mm of the root exposed. Attrition was observed on 99% of teeth examined, however severe dental wear was present in only 5% of the dataset. Both abscesses and enamel hypoplasia are widespread in this study, present in 41% and 26% of the samples respectively. We found no statistical difference between collections in the mean number of teeth exhibiting decay, however we found strong support for significant differences based on root exposure, calculus abundance, and attrition, with a collection of Chinese Americans buried on Kodiak Island, Alaska having the poorest dental health. However, when individuals were grouped by

“race” instead of geography, then there were no statistical differences among black, Asian and white Americans in teeth with caries, antemortem missing teeth and dental wear.

Prior to the third epidemiological transition, which is associated with improved health and life expectancy, dental health of studied American populations was poorer than for people living today. The sample groups examined in this study are comparable to contemporary historic-era populations that largely quantified the dental health of white Americans, and suggest that overall Americans of different racial backgrounds had similar dental health in the early 20th century.

Introduction

Today, oral disease is the most common of all noncommunicable diseases, with dental decay being the most prevalent [10]. Ninety-two percent of US adults between the age of 20-64 have had dental caries and 26% of these adults have untreated dental decay [10]. Nine percent of adults of the same age have periodontal disease, the second most prevalent oral disease. The causes of both these diseases are multifactorial with oral microbial communities, sugar intake, changes in pH, and salivary flow (the amount of liquid produced by the salivary glands) being some of the leading causes [29, 30]. Periodontitis has been associated with systemic diseases as well; this includes stroke [31], cardiovascular disease [32], pneumonia [33], rheumatoid arthritis [34], and Alzheimer’s disease [35]. In addition, oral bacteria, such as *Fusobacterium nucleatum*, have been recovered outside of the oral cavity and is in association with colorectal cancer [36]. Despite the biomedical importance of the study of dental disease, the prevalence of oral diseases over the course of human

history is primarily investigated by paleopathological methods. Their prevalence was not counted systematically in living people in the US until 1959 in the first National Health Examination Survey (NHES) using the decay-missing filled tooth method [8].

The epidemiological transition model provides a framework to evaluate patterns of health and disease as they intersect with demographic, economic, and sociological factors [3, 5]. Omran proposes three successive stages of epidemiological transitions: 1) the Age of Pestilence and Famine, 2) the Age of Receding Pandemics, and 3) the Age of Degenerative and Man-Made Diseases [3, 5]. Each age is marked by shifts in rates of fertility, mortality, and life expectancy [3, 5]. The Age of Pestilence and Famine encapsulates all pre-modern human history and is characterized by high mortality, high fertility, and a life expectancy of 20-40 years [3, 5, 37]. During the second stage, the Age of Receding Pandemics, life expectancy expands to 30-50 years, mortality declines, and fertility remains high causing a sharp increase in population size [3, 5, 37]. Lastly, the third epidemiological transition, the Age of Degenerative and Man-Made Diseases begins following World War II and is associated a further increase in life expectancy and a decline in both fertility and mortality resulting in a slower population growth [3, 5, 37]. This improvement is largely attributed to improvements in diet, advancements in medicine, and economic progress over the first 50 years of the 20th century.

We evaluated the applicability of the epidemiological transition model to human oral health by examining dentition of Americans who lived during the Age of Receding Pandemics (comprised of pre-NHES humans). While systemic infectious and chronic diseases have been assessed through this perspective, the prevalence of oral diseases in the recent past remains understudied. We

provide an important benchmark to examine oral health during the Age of Receding Pandemics. This time period preceded important interventions including dental hygiene and water fluorination, used by most people today, and certainly by 1959, implemented to reduce rates of oral diseases. If the trends in oral disease are the same as systemic diseases as proposed by the epidemiological transition, this might speak to an intrinsic connection between the two. If not, alternative hypotheses should be proposed to investigate why, despite a range of advancements in medicine and dentistry, has oral disease not improved over the last 125 years.

Materials and Methods

The National Museum of Natural History

Medical schools were sources for three of the four populations included in this study. The Huntington Collection, housed by the physical anthropology department at the National Museum of Natural History (NMNH), is populated by unclaimed bodies from the College of Physicians and Surgeons in New York City (referred to as NMNH-Huntington in this study). The collection was originally comprised of 3600 people who died in New York City between 1893 and 1921 and who were subsequently dissected at the college. The collection now consists of about 3070 partial skeletons with ages ranging from 15-96 with birth years between 1798 and 1901 [38]. The collection contains the remains of predominately Irish, German, and Italian immigrants and US-born people recorded as White and Black (“Negro”). There are however individuals from all over the globe, including Asia, the Caribbean, and Australia.

The NMNH also maintains a skeletal collection of individuals autopsied at Howard University in the early 1900s (referred to here as NMNH-Howard). It is

comprised of African Americans who resided in the Washington DC metropolitan area and died in the early 1900s. While little is known about the origins, preservation, and curation of these individuals, each skeletal assemblage is accompanied by autopsy cards with their demographic data.

Specimen preservation procedures impact the availability of skeletal material for anatomists, physical anthropologists, and scientists to study from. Various methods were used to de-flesh skeletons by natural history museums [39]. Curator Aleš Hrdlička notes remains accessioned into anthropology collections at the natural history museum were boiled in steam-piped vats and left to dry on dissecting tables [40]. Prior to being catalogued, remains were often subjected to additional chemical and heat treatments [39].

Kodiak Collection

Our study also includes a population of Chinese migrants from Kodiak Island, Alaska that is archaeological in origin (referred to as NMNH-Kodiak). Skeletal remains from men who worked at Karluk salmon cannery populated a cemetery that was in use from 1882 to the early 1900s and was excavated in 1931 [41]. The immigration of Asian populations to the United States began in mid-1800s with the California Gold Rush [41]. In the 19th century, Chinese immigrants migrated from Guangdong province amid political unrest and settled in the Pacific Northwest and took jobs in mines, factories, and the burgeoning fish cannery business [42]. The incoming Chinese were predominantly male, between the ages of 20-40, with a range of educational, professional, and economic resources [42, 43].

Few archival documents were describe the excavation of this population of salmon cannery workers in late 19th century Alaska. Hrdlička wrote two

sentences in his 486 page book ‘Anthropology of Kodiak Island’ describing the excavation process, “three of the boys excavated in the old Chinese burial ground near the old cannery-found but one body lay 40 odd years in ground, yet much of the tissue still on bones” [44]. The excavation of Kodiak Island gained momentum when Hrdlička formed a collaboration with Laura Jones, the wife of the owner of Alaska Packers Association cannery, in Larsen Bay to transport skeletons and cultural goods from the cemetery to the Smithsonian [45]. Skeletal remains and cultural artifacts were exhumed and transported to the NMNH to be part of Hrdlička’s growing skeletal collection.

The Cleveland Museum of Natural History

Our final sample group comes from the Cleveland Museum of Natural History (CMNH) who accessioned dissected bodies from the former Case Western Reserve University (CMNH-HamannTodd). The Hamann-Todd Collection is comprised of more than 3,000 individuals listed as White, Black (“Negro”), and Chinese (“Oriental”)`, who were born in 1825-1910 and died in Cuyahoga Country, OH or neighboring cities between 1912 and 1938. The collection was started in 1893 by Carl A Hamann, Professor of Anatomy at then Western Reserve University [46]. He was later replaced by T. Wingate Todd in 1912. Details regarding the preservation of the collection are well-documented in the dissertation of Dr. Todd’s graduate student Dr. William Montague Cobb. Upon arrival to the laboratory, cadavers were processed by a prosector who transcribed data from their death certificate, took photographs, and collected additional medical and anthroposcopic data [47]. Cadavers were preserved with embalming fluid and treated with an array of chemicals such as 10% formalin and carmine-red starch prior to dissection [47]. Leonhart’s live steam method

was used to macerate the skeleton over the course of 30 hours [47]. Following maceration, measurements were taken once more to account for shrinkage known to occur following steaming [47, 48]. After maceration, they were cleaned with a rotary wire brush, set out to dry for 2 weeks, then stored and assigned a catalog number [47].

Research Recording Methods

A total of 1123 sockets (from 46 individuals) were evaluated macroscopically from the NMNH and the CMNH and dental pathologies were scored using Hillson (2001) for dental caries, Smith (1984) for dental wear, and Greene et al. (2005) for dental calculus (additional sample information is available in Supplementary Table 2.1) [49-51]. Antemortem and postmortem tooth loss were also noted according to Hillson (2001). Root exposure was measured on the buccal surface with calipers. Abscesses [52] and enamel defects were also documented by individual. Cases of each dental pathology were entered into Excel and imported into RStudio version 3.3.3 [53]. Dental sockets were analyzed by collection, “race”, tooth number, and severity of dental pathologies and results were visualized in ggplot2 version 2.2.1 [54]. Kruskal-Wallis Rank Sum tests were computed in R Studio are reported in the main text. For comparison, ANOVAs were performed in Microsoft Excel version 16.21.1 and are reported in Table 2.5. To facilitate comparison with modern datasets we included an additional index, computed by adding the number of decayed, missing, and filled teeth (DMFT). All measurements and calculations are given in Supplementary Table 2.2.

Results

Summary statistics

This research involved the analysis of a total of 46 individuals (summary table by individual is provided in Table 2.1) and 1123 dental sockets at physical anthropology collections at the NMNH (N=35) and the CMNH (N=11). Measurements ascertained by Hillson (2001), Smith (1984), and Greene et al. (2005) are provided in Supplementary Table 2.2. Of the 1123 dental sockets observed, 689 teeth were present. Additionally we observed gross-gross caries in 37 teeth, 182 teeth were lost antemortem, and 181 teeth were lost postmortem as shown in Table 2.2. Dentitions with occlusal surface gross-gross caries were excluded from all downstream analysis.

Collection	Individuals	Males/Females	Sex Unknown	"White"	"Black"	"Chinese"	"Race" Unknown
NMNH-Huntington	14	7/2	5	7	1	0	6
NMNH-Howard	11	6/3	2	0	11	0	0
NMNH-Kodiak	10	10*	0	0	0	10	0
CMNH-HamannTodd	11	8/3	0	4	5	2	0
Totals	46	31/8	7	11	16	12	6

Table 2.1: Summary sample information by collection. The table includes demographic information also provided in Supplementary Table 2.1. "Race" was assigned during time of death by a pathologist, by medical records of the individual, and/or by the prosector during their accession date.

*The NMNH-Kodiak collection is only comprised of males

Collection	Individuals	Total Sockets	Teeth Present	Occlusal Surface Gross-Gross Caries	AMTL	PMTL
NMNH-Huntington	14	150	51	8	26	56
NMNH-Howard	11	319	200	11	43	61
NMNH-Kodiak	10	320	196	8	63	45
CMNH-HamannTodd	11	334	242	10	50	19
Totals	46	1123	689	37	182	181

Table 2.2: Summary statistics from four skeletal collections under study. This table shows the number of individuals, total number of sockets, total number of observed teeth, occlusal surface gross-gross caries, antemortem tooth loss, and postmortem tooth loss by collection.

Of the collections included in our analysis, the NMNH-Huntington represents the poorest preservation of dentition. From 14 individuals, only 51 teeth were observed. In NMNH-Howard we observed 200 teeth from 11 individuals. The NMNH-Kodiak Collection was the best preserved collection out of all the collections from the NMNH. We observed 196 teeth from 10 individuals in the NMNH-Kodiak. Overall, the CMNH-HamannTodd was the best preserved collection under study with a total of 242 teeth from 11 individuals. All maxillary dental sockets and most of the mandibular dental sockets from each individual were available for our analysis.

Caries

Dental caries were characterized using the scoring sheet published by Hillson (2001). Cavitation on the mesial, distal, buccal, and lingual surfaces was recorded if the site was intact. Across all collections, we measured mesial, distal, buccal, and lingual root and surface caries. Tables 2.2-2.3 shows our counts of small, large, and gross non-occlusal carious lesions of the dental crown and root across all collections. The table includes the total count of each caries type, the number of times the site was measured, and the overall percentage of the caries for each collection in our study.

We observed a total of 115 non-occlusal caries. Thirty-three percent of non-occlusal caries observed in our dataset were from the CMNH-Hamann-Todd, followed by the NMNH-Kodiak (30%) and NMNH-Howard (23%), and lastly the NMNH-Huntington (12%). The distal crown surface had the highest amount of caries observed compared to lingual root caries represented the rarest caries type at 0.4% of all sites observed. We combined gross-gross occlusal caries with all other caries type to calculate the number of decayed teeth per individual

(see Supplementary Table 2.2). It is important to note, that both of these measures exclude the consideration of multiple caries per tooth per collection.

We observed an average of 3.86 decayed teeth per individual in the entire dataset. Across all the collections, the CMNH-HamannTodd had the highest average at 3.72 per person, followed by NMNH-Kodiak at 3.4, and NMNH-Howard at 3.27. The NMNH-Huntington had only 1.4 per person, which may be partially due to the smaller number of teeth available for each individual. Regardless, there were no statistical differences between the means of decayed teeth across the four collections ($p=0.16172$, df (degrees of freedom)=3, $\chi^2=7.3433$).

	NMNH-Huntington		NMNH-Howard		NMNH-Kodiak		CMNH-HamannTodd		Totals	
Distal Root	0%	0/31	0%	0/130	4.8%	4/83	1.2%	2/164	1.4%	6/408
% No./Total										
Distal Surface	11.1%	4/36	6.9%	10/144	8.4%	7/83	10.5%	14/133	9.5%	38/396
% No./Total										
Mesial Root	6.6%	2/30	0%	0/143	4.5%	4/87	0.5%	1/170	1.6%	7/427
% No./Total										
Mesial Surface	15.7%	6/38	7.7%	11/142	10.7%	7/65	7.5%	10/132	9%	34/377
% No./Total										

	NMNH-Huntington		NMNH-Howard		NMNH-Kodiak		CMNH-HamannTodd		Totals	
Lingual Root % No./Total	0%	0/20	0%	0/119	2%	2/100	0%	0/194	0.4%	2/433
Lingual Surface % No./Total	5%	1/20	0.6%	1/146	0%	0/70	1.6%	3/189	1.1%	5/425
Buccal Root % No./Total	5.8%	1/17	0.0%	1/108	6.3%	7/111	3.8%	7/182	3.8%	16/418
Buccal Surface % No./Total	0%	0/16	3%	4/131	0.8%	1/122	1%	2/191	1.5%	7/460

Tables 2.3 and 2.4: By collection, the scores for caries given according to Hillson (2001) are shown by percentages evaluated in tooth as well as the overall number and the total sites observed.

Periodontitis

While specific and consistently used definitions distinguishing types of periodontitis are unavailable for paleopathologists, we measured the tooth from the cement-enamel junction (CEJ) to the alveolar bone lining socket to determine the extent of attachment as a measure of periodontitis [49]. The buccal surface was the most accessible surface to measure root exposure across all collections. Measurements were taken to the nearest millimeter from the cement-enamel junction (CEJ) to the alveolar bone lining [49]. Measurements were not taken if there was evidence of post-mortem damage at the alveolar process, the CEJ was difficult to locate, calculus covered the buccal surface, or caries destroyed the site altogether.

We measured these buccal root exposures on 540/689 teeth. Only a single lower incisor had 0 mm of the root exposed. All others ranged from 1-20 mm as shown in Figure 2.1. The largest exposure was on a single individual's maxillary canine in the NMNH-Kodiak, which measured 20 mm. The average buccal root exposure for dentition across all collections was highest in the NMNH-Kodiak at 5.40, followed by the NMNH-Huntington at 4.06, the NMNH-Howard at 4.01, and lastly the CMNH-HamannTodd 3.84. In contrast to caries, we found statistically significant differences between the collections for buccal root exposure ($p=7.715E-10$, $df=3$, $\chi^2=45.371$).

We also evaluated these ranges on an individual-by-individual basis. We found that a majority of individuals (74%; 34/46) also had at least one tooth with 3-4mm of the buccal root exposed. This included all of the individuals within the CMNH-HamannTodd collection and the NMNH-Kodiak collection. Root exposure up to and exceeding 5 mm was seen in 69% of individuals in the dataset, including all of the individuals within the NMNH-Kodiak. The least observed root

exposure range was 1-2mm across all collections, however observed in 54% (25/46) of individuals in this study. Additionally, periodontitis is known to increase with age (CDC 2018). We assessed the severity of root exposure with the few individuals whose age at death was recorded. Root exposure >3 mm was observed in all individuals with known age (Supplementary Table 2.2).

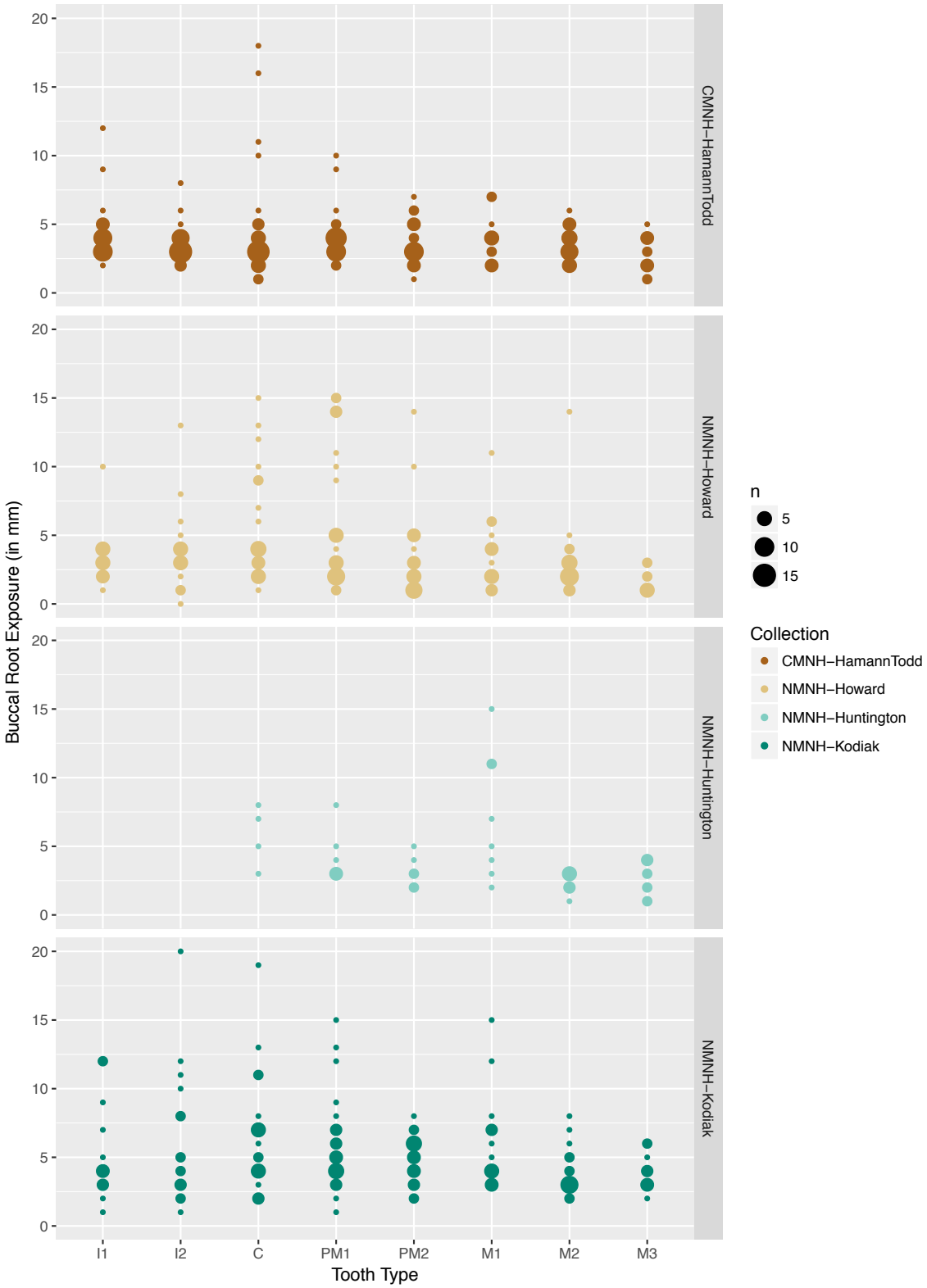


Figure 2.1: Buccal root exposure (in mm) by tooth type across all collections. I1 and I2 are the first and second incisors. C stands for canine, PM1 and PM2 are the first and second premolar. M1, M2, and M3 are the first, second, and third molars.

Dental calculus

Figures 2.2A-C illustrate the distribution of calculus across all collections according to Greene et al. 2005. Calculus was measured on the interproximal, lingual, and labial/buccal surfaces of 2016/2067 surfaces available for our analysis shown in Supplementary Table 2.3. Out of 2016 dental surfaces observed, no calculus was found in nearly half (48%; 972/2016) of the sites analyzed. Minimal amounts, covering about 1/3 of the surface, was observed 31% (626/2016), moderate amounts 10% (211/2016), and extensive amounts, covering more than about 2/3 of the surface, 13% (256/2016) of the time. Of the surfaces evaluated, interproximal calculus accounted for a majority of the sites with minimal-extensive amounts of calculus (20%; 410/2016), followed by lingual (17%, 338/2016), and lastly labial/buccal (14%; 296/2016).

Of the sites with extensive calculus, most is found on the labial/buccal surface (38%; 98/256), followed by the lingual surface (32%; 81/256), and lastly the interproximal surface (30%; 77/256). When we assessed the prevalence of extensive calculus across collections we found that NMNH-Kodiak has the highest amount of tooth sites with extensive calculus, given the score of 3 in Greene et al. (2005), (40%; 104/256), in contrast to the NMNH-Huntington where extensive calculus was least prevalent (14%; 36/256).

We also evaluated statistical differences between the scores assigned to each site in each collection. Similar to root exposure, we found statistical significant differences between collections in all sites measured: labial/buccal ($p= 1.528E-07$, $df=3$, $\chi^2=34.534$); lingual ($p=0.55844$, $df=3$, $\chi^2=7.64658$); interproximal ($p= 1.982E-08$, $df=3$, $\chi^2=38.728$). NMNH-Huntington, despite contributing the fewest teeth to our analysis, had the highest averages for both

labial/buccal and interproximal calculus. NMNH-Kodiak had the highest average for lingual calculus.

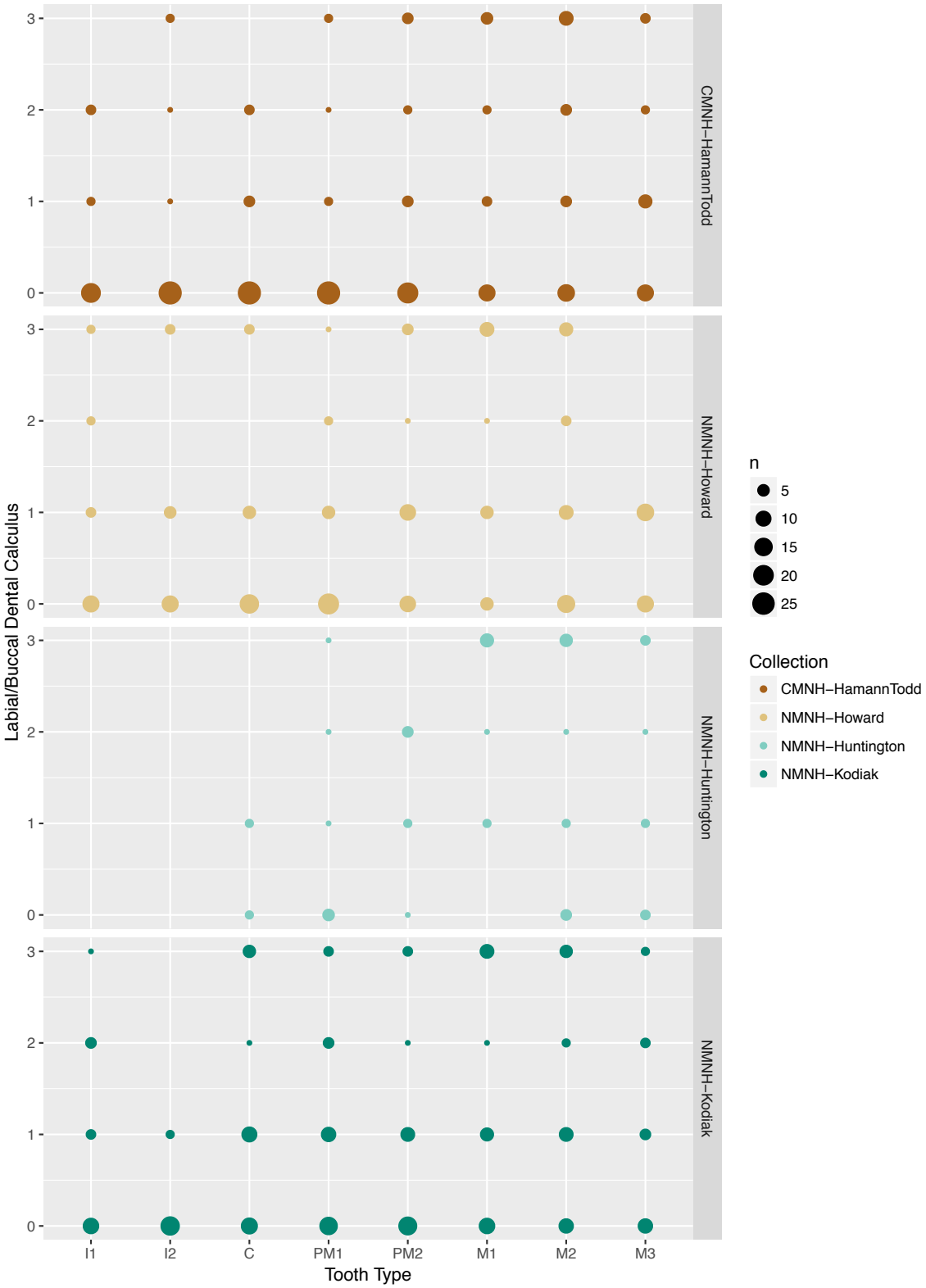
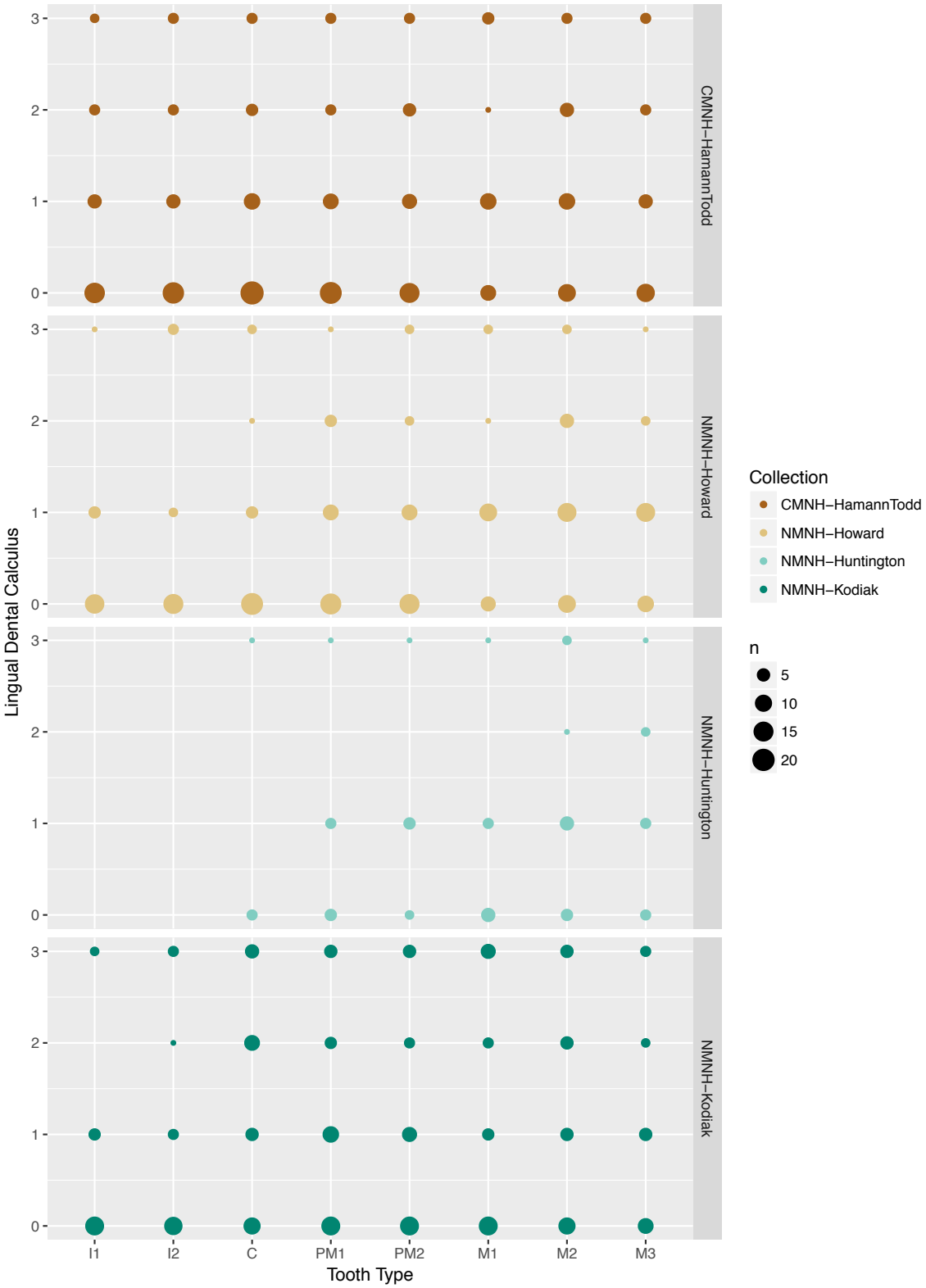


Figure 2.2A: Interproximal calculus abundance by tooth type all across collections. I1 and I2 are the first and second incisors. C stands for canine, PM1 and PM2 are the first and second premolar. M1, M2, and M3 are the first, second, and third molars.



Figures 2.2B: Lingual calculus abundance by tooth type all across collections. I1 and I2 are the first and second incisors. C stands for canine, PM1 and PM2 are the first and second premolar. M1, M2, and M3 are the first, second, and third molars.

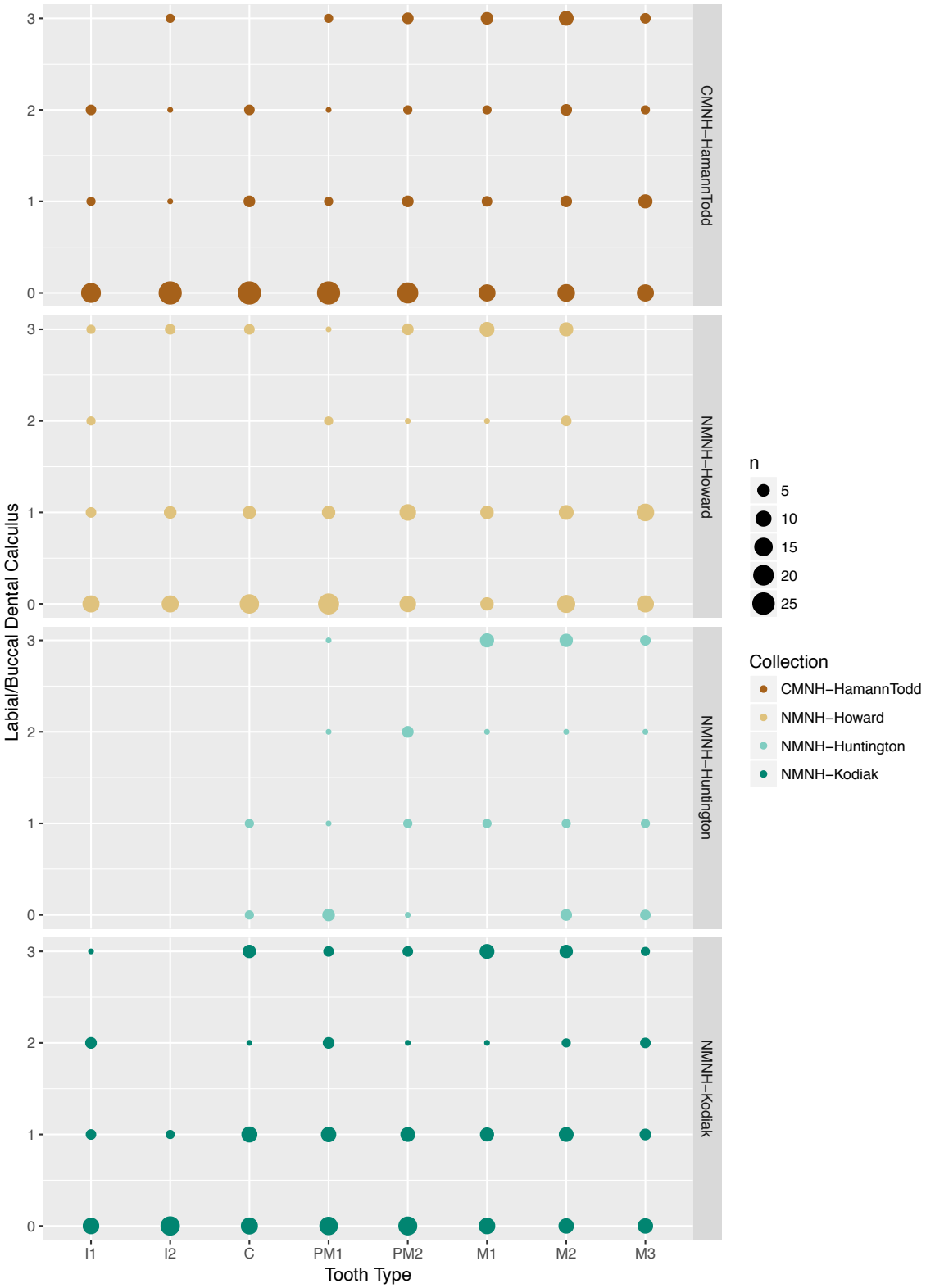


Figure 2.2C: Labial/buccal calculus abundance by tooth type all across collections. I1 and I2 are the first and second incisors. C stands for canine, PM1 and PM2 are the first and second premolar. M1, M2, and M3 are the first, second, and third molars.

Other Pathologies

Dentition was also evaluated for wear. The severity of attrition was examined on 629 teeth using methods described by Smith 1984 shown in Supplementary Table 2.4. Teeth were not scored if the occlusal surface was destroyed by caries, covered in calculus, or stained. Of the dentition available for our analysis, only 3 teeth had no signs of dental wear. The remaining dentition (99%; 625/629) showed mild, moderate, and severe signs of wear. For all collections, mild attrition counted for 58-75% of sites observed. Moderate attrition, which received scores of 3 and 4 in Smith 1984, ranged from 20-31% of each of the collections. Finally severe attrition, resulting in large dentin exposures and the loss of crown height and surface, accounted for 2-9% of the teeth observed for each collection in our analysis.

The most severe wear stages were observed in the NMNH-Kodiak. Dental wear was extensive across 4 individuals of the NMNH-Kodiak. Interestingly, in individual 364459 wear was observed on 3 of their first molars. Individual 364438 had severe dental wear on their upper first and second right molars. Individual 364456 had dental wear localized on their upper and lower left canines and their lower right first molar. Individual 364455 had severe dental attrition in their upper left canine and first molar and lower right incisor and canine. The most severe wear observed on a single individual HTH399 with 7 teeth exhibiting large dentin exposures, coalescence of dentinal areas with the enamel rim still intact. Two other individuals from CMNH-HamannTodd had extensive wear patterns: HTH249 on their right I1 and left I2 and HTH475 their right canine.

We tested for statistical differences between collections based on the attrition score of the dentition and the average wear score of individuals within

each collection. Dentition varied among the collections (p=0.0001419, df=3, χ^2=30.276). However, when a mean wear score is calculated by individual and the collections are compared, there are no statistically significant differences among the collections (p-value=0.2277, df=3, χ^2=4.3332).

The presence of abscesses and enamel defects were evaluated by individual per collection. Abscesses were widespread across all collections, present in individuals of all collections analyzed (41%; 19/46). Probable enamel defects were present in 26% (12/46) of all individuals included in our study. Two of these cases are seen in Figures 2.3A and 2.3B.

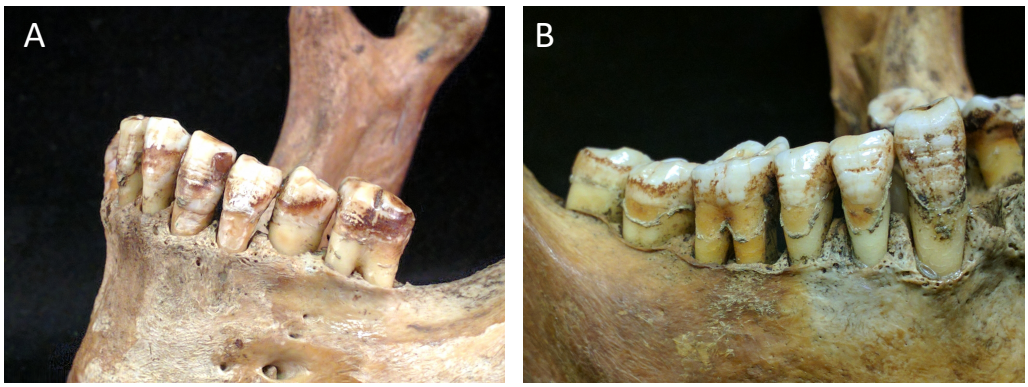


Figure 3A and 3B: Images of mandibles from individuals 364456 and 364459 respectively. Enamel defects, calculus, and buccal root exposure are seen above.

Given the breadth of dental paleopathology data collected, we tested for associations between the different pathologies analyzed – non-occlusal caries, root exposure, calculus distribution, and attrition and found no strong positive or negative associations when all of the pathologies are considered together.

DMFT Index

We calculated this score based on the presence of fillings in individuals within our sample. The minimum DMFT is 0 and our maximum is 28. We found four individuals with dental fillings: 2 individuals in the NMNH-Hamann-Todd, 1 in the NMNH-Huntington, and 1 in the NMNH-Howard. The average DMFT across all collections was 6.8, indicating each individual on average had an average of six teeth that were decayed, missing, or filled.

Comparison between "races"

While two of the four collections studied are comprised of solely a single "race" (the NMNH-Kodiak are Chinese immigrants; the NMNH-Howard are black Americans), the remaining two collections are racially mixed. When individuals are separated by "race" into black Americans (n=17), Chinese Americans (n=12) and white Americans (n=11), we found that all three groups had similar levels of measured dental pathologies. Per individual, the number of teeth with caries ($p=0.4511$, $df=2$, $\chi^2=1.5923$), antemortem missing teeth ($p=.3085$, $df=2$, $\chi^2=6.9573$) and dental wear ($p=0.1578$, $df=2$, $\chi^2=3.6927$) were similar across all three groups. The number of antemortem missing teeth was qualitatively lower in white Americans than in the other racial groups, but none of these differences reached statistical significance. Combined with the observation that the four individuals with fillings were from diverse backgrounds (two were black, one Asian and one white), this provides evidence that dental health was similar across racial groups under study.

Statistical Tests	Kruskal-Wallis Test			ANOVA		
	p-value	X^2	df	p-value	F	df
Number of decayed teeth by collection	0.16172	7.3433	3	0.217	1.545	3
Buccal root exposure by collection	7.715E-10	45.371	3	9.0405E-06	8.91	3
Labial/buccal dental calculus by collection	1.528E-07	34.534	3	4.858E-07	10.971	3
Lingual dental calculus by collection	0.55844	7.64658	3	0.0085	3.928	3
Interproximal dental calculus by collection	1.982E-08	38.728	3	1.125E-09	15.369	3
Attrition by dentition across collections	0.0001419	30.276	3	0.0003	6.2236	3
Mean attrition of individuals across collections	0.2277	4.3332	3	0.504	0.7944	3
Average number of decayed teeth by race	0.4511	1.5923	2	0.656	0.426	2
Average number of missing teeth by race	0.3085	6.9573	2	0.103	2.424	2
Mean attrition scores by race	0.1578	3.6927	2	0.402	0.933	2

Table 2.5: Results of statistical tests

Discussion

In this study, we sought to investigate the state of dental health at the turn of the century and the suitability of the epidemiological transition theory in describing the natural history of dental diseases. Depending on techniques used following specimen acquisition and specimen preservation, dentition is differentially abundant across sampled collections. We expected a total of 1,472 total sockets to be preserved for our analysis, however, between 6-37% of dentition was lost postmortem. The most complete dental arcades we examined were housed by the CMNH-HamannTodd with only 6% of dentition lost postmortem. This was possibly due to the great care and attention of skull by curators at the CMNH. Each skull was cleaned and scraped separately from the rest of the skeleton [47]. Several measurements were taken across the craniofacial region and cranial capacity measurements required the complete crania to ascertain [47]. Even the smallest, most fragile bones of the crania remained intact for special study [47].

The poorest preserved collection was NMNH-Huntington with 37% of dentition lost postmortem. This is likely due to the nature of the collection's establishment and its transport from New York City to the Smithsonian in Washington, D.C. The collection was originally established by morphologist Dr. George Sumner Huntington, who collected 500 of each bone of the human body [38]. It was transferred to Aleš Hrdlička's care over many years, resulting in lost documentation and remains, and resulting in commingling [55].

While cultural, environmental, curation, and other processes obstructed our ability to evaluate all dentition available for all pathologies, we also observed many damaged sockets that were destroyed postmortem [39]. Chipping was observed in nearly a third of all individuals in our dataset. Museum

specimens are handled by many physical anthropologists, anatomists, and other researchers who may mishandle and drop remains on occasion. In addition, glue was observed on 23% of individuals on the mesial, distal, and/or root surfaces of dentition. Some specimens were not completely macerated and flesh remained on the bone and interproximal regions of teeth as well. Also calculus was predominantly abundant on the interproximal surface, which made it difficult to score mesial and distal caries on teeth. We also found indications of smoking across all collections (50% of the samples), however most frequently in the NMNH-Kodiak collection. This is consistent with the account of life of the Chinese immigrants in Alaska who commonly smoked opium to cope with the stresses of cannery living [43].

Skeletal collections are typically made up of males [56], and only 17% of individuals in our sample were identified as female. Additionally, age was not estimated using postcranial indicators, but rather only by the presence of the erupted of the third molar. Ages are known for 39% of the sample, provided by medical records. They range from 22-70 and with an average of 37.5. Due to the absence of demographic data, we did not have the statistical power to estimate statistically significant differences between individuals based on age or sex. Further evaluation of these collections may provide insight into historical dental health disparities by these demographic factors.

Caries

Dental caries are the result of destruction of the dental enamel by the release of acid from resident microflora [16, 57]. Bacteria, such as mutans streptococci, ferment dietary carbohydrates and release lactic acid into the mouth [16, 57]. While the development of caries is multifactorial, diet,

specifically, the consumption of sugar, plays a major role [57]. During the turn of the 20th century, there was also marked increase in sugar consumption [57, 58]. Other foods frequently eaten at the turn of the century can also cause dental disease. Potato starch and lactose can reduce to the pH of the oral cavity and has been shown to contribute to the development of caries in a mid-nineteenth century Irish population [59, 60].

The prevalence rate of caries was 22.1%, which is comparable to contemporaries from US, Canada, and Europe [61-63] (Supplementary Table 2.5). Despite possibly sharing similar socioeconomic statuses, the prevalence of carious teeth in our study is lower than a sample of impoverished Irish workhouse laborers who died between 1847-1851 [59]. This trend is also found when comparing the skeletal remains from a wealthy family plantation owners who died between 1830s-1907 [63]. Most of the contemporary studies are solely comprised of Europeans or white Americans (one notable exception is the Golden Gate Cemetery excavation, which likely included white, Chinese and Native Americans) [62]. That the number of decayed teeth of our racially diverse Americans is similar to that of contemporary white Americans is further evidence for the general dental health of all Americans being similar in the early 20th century. Nearly all studies of North American and European dental pathology from this time period have higher prevalence of dental caries and antemortem tooth loss than that of contemporary Khoesan individuals [64], which shows that the more common environments and diets of early 20th century western life lead to poorer dental health. Further examination of diet through stable isotopic and proteomics analysis may help to identify dietary components that influence the development and severity of caries.

Dental calculus

Dental calculus is mineralized dental plaque. High calculus formers have been shown to also have high concentrations of calcium and phosphate in their saliva [65]. It develops during life and requires removal using specialized dental instruments [16]. In the case of calculus, prevalence is comparable to the agriculturalists described in Botha and Steyn (2015). However, it was not as ubiquitous as observed in a population of mid-19th century poor Irish workhouse laborers [59]. As evidenced by the presence of fillings, the study population described in our study may have had more affordable and frequent access to oral care and hygiene products than their contemporaries.

Dental medicine and oral hygiene

Dental hygiene, water fluoridation, and regular dental visits have contributed to the reduction of dental disease over the last century. In 1908, the American Dental Association (ADA) issued its first dental education pamphlet for patient [66]. The pamphlet established modern day dental hygiene recommendations by advocating that their patients brush their teeth two a day, floss regularly, and visit the dentist office twice a year [66]. These recommendations are found on prominent dental products of the time: toothpastes, brushes, and washes produced by Colgate, Listerine, and others. Records regarding access to a dentist and frequency of visits are impossible to detect from the skeleton and are not documented in medical records. While it is largely believed that the skeletal remains housed in physical anthropology collections at natural history museums reflect the poorest and most destitute people at the time, we provide the evidence of access to dentists also seen in contemporaneous studies [61, 63, 67, 68]. The average number of fillings found

in our study is 0.08 compared to 4.61-9.20 of adults today [69]. While dental fillings fluctuated in affordability (and safety) prior and during the “Amalgam Wars,” fillings were relatively expensive until after WWII [70]. None of the individuals within the NMNH-Kodiak sample had fillings – and these individuals had the most significant evidence of periodontal disease, the most calculus and the greatest attrition. We suspect this comparative lack of dental care is due to the transient and seasonal nature of labor in canneries in the 1800 and early 1900s.

Disease Indices

Diseases indices ensure the comparability of working datasets to those that were published. We calculated the DMFT for every individual in our study based on the presence of fillings. The DMFT and the DMFS (a calculation analogous to DMFT for each surface of a tooth) are standard methods for estimating overall dental health in people since 1930s [71, 72]. However, the usage of the DMFT score in paleopathological research is limited due to the unknown etiology of missing teeth. A range of infectious diseases, such as caries, periodontitis, and abscesses, can lead to edentulousness. The “missing” component of the entire DMFT is also contested within the dental medicine literature because it can lead to the under or overestimation of caries experience [73]. However, despite these limitations, the average decayed dentition is higher than modern datasets, but the average number of filled teeth is lower than modern people living today as shown in this study [69].

To assess the suitability of the epidemiological transition theory, we must investigate dental health over the 20th century and design new disease indices to do so given the accessibility to dentistry, dental hygiene, and water

fluorination unavailable to people before the 20th century. Other indices used in dental medicine do not suffice our purposes given that paleopathologists work with skeletal remains. For example, the NHANES uses the simplified oral hygiene index, which may be useful given the increased usage and affordability of dental hygiene products over the 20th century. However, estimating debris in skeletal remains is limited due to maceration practices. Additionally, the presence of calculus can greatly vary in historic-era individuals, ranging from 44-97.3% of late 19th and early 20th century sample groups [59, 64].

The osteological paradox prevents bioarchaeologists from using disease frequencies in skeletal assemblages as a representation of their frequencies amongst the living population [74]. Proposed by Bernard Wood and colleagues (1992), the osteological paradox forced bioarchaeologists to address three conceptual problems while performing statistical analyses and inferring health from past populations: demographic nonstationarity, selective mortality, and hidden heterogeneity in frailty [74]. Given the historical context of these collections, we cannot view these samples as representative of the overall populations from which they were drawn [56]. This study was undertaken on skeletal collections comprised of unclaimed bodies and immigrant working class, likely from the poorest classes of society. We should not assume all segments of the population went through the same demographic transitions concurrently [3, 75, 76].

While “race” does not have a biological basis, we now know that “race” does in fact have very real consequences on the lived experiences of people, and in this way, “race becomes biology” [77]. Social segregation, based on socioeconomic status and/or “race”, puts marginalized people into low income neighborhoods, with inadequate housing, and limited access to healthcare [78].

This intersection of “race” and socioeconomic status can result in a pattern where the poorest segments of the population are exposed to infectious diseases, while the affluent classes experience an increased in chronic non-infectious diseases [76]. We should also note that currently dental services utilization inequalities, in fact, are worldwide, with ethnic minorities and those with low education, income, and poor insurance coverage having consistently lower access [69, 79].

Additionally, while dental diseases are typically chronic in nature, they have varying effects on an individual’s frailty. For example, calculus accumulation is associated with high frailty and an elevated risk of mortality in medieval London [80]. If other diseases are considered in conjunction with demographic information, such as mortality and life expectancy, it may provide stronger support or present a challenge to the epidemiological transition theory [80].

Given the osteological paradox, it may be difficult to reconstruct the historical trajectory of dental disease within the United States. Our study found that late 19th and early 20th century Americans, had more decayed teeth and less fillings by using the DMFT. The epidemiological transition theory suggests chronic disease have increased in prevalence, while infectious diseases have decreased for the general population [5]. If we view dental diseases, such as dental calculus abundance, as chronic in nature, the trajectory of dental disease in the US does not coincide with systemic diseases. However given the etiology of dental diseases such as dental caries, which may be infectious in nature, which have decreased over time, this supports the epidemiological transition theory [81]. It may be possible for the epidemiological transition theory to explain trends in certain dental diseases, yet contradict others, we believe

additional research in historical dental pathologies are needed to address this question.

Conclusion

We analyzed 1123 dental sockets from 46 individuals from geographically and racially diverse backgrounds. Using a range of methods available, we can tentatively state that prior to the third epidemiological transition, dental health was poorer than that of present day Americans, but comparable to contemporary historic-era populations. We found statistically significant differences for root exposure, calculus abundance, and attrition between sample groups, each skewed towards more severe dental pathologies in the NMNH-Kodiak collection, but no difference in the number of non-occlusal caries. Despite the differences observed between collections determined by geography, we found that dental pathologies were similarly prevalent across black, Asian and white Americans. Altogether, this snapshot of diverse Americans provide an important comparative benchmark and further examination will shed light on the history of oral health, historical health disparities, and the impact of medical and dental health interventions over the course of the 20th century.

CHAPTER 3: THE "MICROBIAL MELTING POT": CHARACTERIZING THE DENTAL CALCULUS MICROBIOME OF TURN OF THE CENTURY AMERICANS

Abstract

The oral microbiome contributes to disease in over 90% of humans, and is one of the most intensively studied microbiomes: seven of the 15 sites sampled in the human microbiome project are in the mouth. It is possible to study historical oral microbiomes because dental calculus (tartar) that forms on humans in life also preserves biomolecules within its matrix for hundreds to thousands of years. While most ancient microbiome research has focused on ancient samples, and overwhelmingly the individuals sampled have been European, here we study the oral microbiome of those who lived just before the technological and socio-cultural changes that define modern American life. We analyzed dental calculus metagenomes from geographically and racially diverse skeletal remains (from Chinese migrants and black and white Americans) who died in the USA between 1895-1950. These showed similar levels of microbial beta diversity, despite differences in sex, "race", and geographic origin. Most of the communities (41/43) have significant archaeal components, up to 74.35% of mapped reads, when archaea are very rare in modern oral microbiomes. Complete high-quality genomes of the archaeon *Methanobrevibacter oralis* were obtained from three individuals (60-99x average coverage), and the two sampled communities without detectable Archaea had sufficient reads to reconstruct two modern nosocomial pathogens (*Stentrophomonas maltophilia* (17x average coverage) and *Acinetobacter baumannii* (7.8x average coverage). The oral microbiome of Americans has dramatically changed in the last century, and since a majority of Americans still suffer from microbial-associated oral disease, these changes are not necessarily associated with

improved oral health. Understanding the makeup of recent historical microbiomes should inform the medical use and modification of human-associated microbiomes.

Introduction

The turn of the 20th century marked a major transition in human evolution. Political, social, cultural, and population change during this time has had lasting effects on human health, diet, and the environment. These effects have been identified from the macroscopic to the microscopic level, potentially causing shifts on the microbes that reside in and our bodies, collectively known as a microbiome [22, 82]. Our microbiomes play major roles in maintaining health and perturbations in the microbiota can cause disease [83]. Researchers suspect western lifestyles and advances in medicine may have limited exposure and influence the inheritance of our old “microbial friends”, influencing our risk for a number of systemic diseases today [4, 84, 85].

While many human-associated microbiomes are being explored, only one microbiome is known to cause disease in a majority of humans today: the oral microbiome [86, 87]. In the US, 91% of adults between the ages of 20-64 have dental caries and 44.7% of adults over 30 have periodontitis [86, 87]. In health, oral microorganisms perform a wide range of functions related to digestion and degradation of food, while also enhancing host immune function, via colonization resistance and the production of antimicrobial compounds [88]. Dental diseases are the result of major ecological disruptions, such as a change in diet that alters available oxygen, pH levels, and microbial populations in the mouth [29].

Altered oral flora have also been linked to extra-oral diseases including obesity, diabetes, bacteremia, preterm birth, and low birth weight in infants [15]. In addition, recent work has investigated the influence of oral bacteria on the

risk of developing several cancers such as oral, gastric, and pancreatic cancer[89-91]. Oral microbiota can leave the oral cavity through the gingival crevice and enter the bloodstream, reaching vital organs throughout the body. For example, lactobacilli and viridans streptococci are associated with infective endocarditis-related bacteremia [92], periodontal pathogen *Fusobacterium nucleatum* has been recovered in colorectal carcinomas [93] and *Porphyromonas gingivalis* has been isolated from the brains of people with Alzheimer's disease [94].

This growing body of research has expanded to provide crucial baselines for the role of agriculture, industry, and colonialism on the oral microbiota [22, 27, 82]. However previous examinations of the evolution of the oral microbiome face several limitations related to sample size, host diversity, and fail to capture recent human history. As we attempt to describe the 'healthy' oral microbiome, there is a need for transitional historical data to provide context to modern oral microbial ecology and evolution and determine if and how it had changed in response to the last 125 years of rapid sociocultural change in diet, improved sanitation, and medical technology [95]. In this study, we characterize the dental calculus metagenomes of geographically and racially diverse Americans who died between 1890-1945.

Materials and Methods

Historical Dental Calculus Sampling

Dental calculus samples were obtained through four geographic and racially diverse anthropology collections maintained at the National Museum of Natural History (NMNH; n=21; 1895-1920), the Cleveland Museum of Natural History (CMNH; n=8; 1912-1938), and the American Museum of Natural History

(AMNH; n=8; 1945-1947; Sample information available in Supplementary Table 3.1). Destructive sample protocols were submitted and approved by physical anthropology sampling committees at the NMNH (1/25/2017), CMNH (11/3/2016), and the AMNH (1/12/2017). The details of how these collections were obtained and preserved are described in Chapter 2. Prior to sampling, the remains were evaluated for dental pathologies using tools developed by Buikstra and Ubelaker 1994, Hillson 2001, and Smith 1984. Results from dental paleopathology are described in Chapter 2 and results are shown in Supplementary Table 2.2.

Prior to calculus removal, a photograph of the remains was taken. During calculus collection, nitrile gloves, a hairnet, and facemask was worn. Dental calculus was isolated using a dental scaler (from the Omni Dental 6 Piece Dentist Tool Kit) over aluminum foil (to catch the fragments as they are removed) on a sterilized table in research space of the physical anthropology collections in the respective museums. After collected, the sample was poured into a sterile 1.5 ml tube, labeled with sample information and the date. Bleach (Clorox bleach germicidal wipes) and/or 70% isopropanol wipes were used to clean the scaler after each sample.

Sample Decontamination and Decalcification

Calculus samples were analyzed in a dedicated ancient DNA laboratory, the Laboratories of Molecular Anthropology and Microbiome Research, at the University of Oklahoma. Photographs were taken of samples in 1.5 ml tubes on arrival. Samples were also weighed and subsampled for further analysis. Each sample was decontaminated with UV light for 1 minute on each side. Samples were washed with l ml of 0.5M EDTA, placed on a nutator for 15 minutes, and

centrifuged at 13K rpm for 3 minutes. The supernatant was then transferred to a new tube and 1 ml of 0.5M EDTA was added to the pellet and placed on the nutator for 24 hours. After 24 hours, 100ul of Qiagen Proteinase K was added to each sample. Samples were placed back on the nutator for further decalcification for 48-72 hours.

DNA extraction and quantification

Following decalcification, all samples were centrifuged and the supernatant was added to 13 ml of PB buffer from the Qiagen PCR Purification MinElute kit. The MinElute column was spun down for 4 minutes at 15000 g, rotated 90 degrees, and spun for another 2 more minutes. 700ul of PE buffer was added to the samples and spun for 10000 rpm. Waste was removed from tubes and an additional 700 ul was added to samples and spun for another minute at 10000 rpm. Waste was removed and tubes were spun one last time at 13000 rpm for a minute. 15 ul of EB buffer was added twice directly to the membrane of MinElute tubes and spun down for 1 minute at 13000 rpm. After extraction, DNA was quantified using a Qubit fluorometer according to protocol. Extraction and library negatives were included during the processing of each batch and sequenced [96].

Shotgun library preparation

Using a NEBNext kit, DNA ends were repaired using 5 ul of buffer, 2.5 ul of end repair enzyme, and 42.5 ul of DNA. The mixture was incubated for 40 minutes at room temperature and for 20 minutes at 37C. This step was followed by the modified MinElute clean up protocol in “DNA extraction and quantification”. Adaptor ligation was completed with 10 ul of quick ligation

buffer, 3 ul of adaptor mix, 3 ul of water, 30 ul of end-repaired sample, and 5 ul of quick T4 ligase from NEBNext kit. The mixture was incubated for 20 minutes at room temperature and washed following the modified MinElute clean up protocol described in “DNA extraction and quantification.” Adaptor fill-in was completed with 5 ul of reaction buffer, 2 ul of Bst polymerase, 13 ul of H20, and 30 ul of ligated sample. This mixture was incubated for 30 minutes at 37C and 20 minutes at 80C. To determine the optimal number of cycles for indexing, we ran a test QPCR using 1 ul of the sample, 5 ul of water, 12.5 ul of Kapa HiFi + Uracil Mix, 1 ul of 2.5 mg/ml BSA, 0.75 ul of forward primer, 0.75ul of reverse primer, 4 u of 50 uM SYTO9 Green. Each of samples were run for 55 cycles: 5 minutes at 94 for initial denaturing, 20 seconds at 98 for denaturing, 15 seconds at 50 for annealing, and 30 seconds at 72 for extension.

CT values were recorded and used to determine cycle numbers for indexing PCR. We used 4 ul of our adaptor fill-in in addition to 13.25 ul of water, 5 ul of Phusion HF Buffer, 0.5ul 2.5 mg/ml BSA, 0.5ul of dnTPs, 0.75ul of primer F, 0.75ul of primer R and 0.25ul of Phusion HS enzyme. Each cluster was run for 30 seconds at 98 for initial denaturing, ten seconds at 98 for denaturing, fifteen seconds at 60 for annealing, and thirty seconds at 72 for elongation. This was followed by the modified MinElute clean up protocol described in “DNA extraction and quantification.”

To determine the length of indexed sequences and volume for shotgun sequencing, the Agilent High Sensitivity DNA Kit was used according to protocols. Size distribution was visualized using a Bioanalyzer and libraries were pooled based on DNA quantities. To remove DNA shorter than 160 base pairs (which were likely primer dimers), we used Pippin Prep. Libraries were sequenced using 2x150 paired end on Illumina HiSeq and NextSeq at Yale

Center for Genome Analysis and Oklahoma Medical Research Foundation Clinical Genomics Center. Samples were sequenced at a depth of approximately 10 million reads. A detailed flow chart of all wet laboratory work is available in Supplementary Figure 3.2.

Computational Analysis

Sequencing quality was assessed using FASTQC version 0.11.8. Sequencing adapters were removed, reads were merged, and quality filtered in AdapterRemoval version 2.1.7 [97]. Quality filtered reads were imported into metaphlan2 for taxonomic assignments using default parameters [98]. To authenticate the historical origin of sequences, reads were mapped to the human genome (hg18), human mitochondrial genome, and 20 of the prevalent microbial species assigned by metaphlan2 using bwa aln -l 1000 -n 0.01 -t 8 [99]. Resulting sam files were converted into bam files and consequently filtered, sorted, and duplicates were removed in samtools [100]. DNA damage patterns were assessed using mapdamage 2.0 on the quality filtered bam files [101]. In addition, potential contamination sources were tracked using QIIME and sourcetracker v1.0.1 [102, 103].

We used a read and assembly based approach for determining taxonomic profiles of historical calculus samples. First, contaminant analysis and removal of human DNA was conducted in kneaddata (https://bitbucket.org/biobakery/kneaddata/wiki/Home). Following decontamination, reads were aligned using DIAMOND to bacterial and archaeal reference protein sequences and all viral proteins available on GenBank (downloaded 3/2019, 70,800,065 amino acid sequences) [104]. Both decontaminated sequences and resulting DIAMOND text files were used to

create RMA files for analysis in MEGAN [105]. Protein accession, VFDB, and SEED mapping files were used for taxonomic, functional, and virulence assignments respectively. Species-level taxonomic assignments were exported for additional decontamination analysis using the decontam package in R [106]. Thresholds were set to 0.1 and 0.5 and resulted in no species being removed from the dataset. For comparative purposes, 16 supragingival and subgingival plaque samples (Supplementary Table 3.3) were downloaded from the HMP Browser, merged in PEAR [107], decontaminated in kneaddata, and analyzed in tandem in MEGAN [105]. Additional ancient [27] and historical [108] dental calculus metagenomic data was downloaded using the Galaxy "Download and Extract Reads in FASTA/FASTQ function (https://usegalaxy.org/) and run through the same bioinformatics pipeline as the this study's samples and the HMP list with SRA number information is available in Supplementary Table 3.4). A secondary metagenomic analysis was completed in metaphlan2 with decontaminated reads [98].

For assembly-based analyses, decontaminated reads were assembled in megahit v.2.2.1 [109], quality checked in quast [110], and uploaded to PATRIC 3.5.31 for metagenomics binning [111]. Reads were mapped back to the PATRIC reference using bowtie2 and overall alignment rates from these alignments are reported in Table 1. Contigs from high quality bins (at least 80% coarse consistency, at least 80% fine consistency, less than 10% contamination, and a single PheS protein) were reordered in Mauve [112], uploaded to the RAST server for annotations [113], and visualized in BRIG [114]. Additional analysis of generated bins was conducted in PanPhlan [115] for strain tracking and function analysis, snippy for variant calling (https://github.com/tseemann/snippy), and PHAST [116] for prophage discovery.

Nucleotide sequences of select gene family clusters underwent BLASTX analysis to locate homologous protein sequences.

All graphic images were created in RStudio using ComplexHeatmaps and ggplot2 [53, 54, 117].

Statistical Analysis

Species level taxonomy tables were converted into presence-absence tables and analyzed in R package indicspecies for indicator species analyses [118]. Species with p-values <0.005 were included in downstream analysis. Presence-absence tables also underwent co-occurrence analysis in R package cooccur [119]. We report all associations made above the threshold and their probabilities to cooccur across samples.

A detailed overview of the bioinformatics pipeline used in this chapter is available in Supplementary Figure 3.5.

Results

Assessing DNA quality and contamination of sequences

A total of 44 samples and 12 blanks were analyzed in this study. DNA yields from dental calculus ranged from 1.3ng/mg – 63.6ng/mg (median 1.99ng/mg). One of the 44 samples failed to sequence, leaving 329,372,806 reads from total shotgun sequencing for the 43 samples and 2,245,236 reads for 12 blanks. Following quality filtering and trimming, a total of 225,772,421 reads were available for downstream analysis. We used Metaphlan2 to determine the absolute abundance of microbial communities for DNA damage authentication. The list of prokaryotes our reads were mapped to for mapdamage analysis are in Supplementary Table 3.6. Despite the use of a proofreading enzyme for PCR,

we observed characteristic patterns of DNA damage shown across all domains of life, including viruses, in Supplementary Figure 3.7.

Additionally, potential contamination within the calculus samples was assessed further through Sourcetracker analysis by analyzing human skin[120], human plaque[121], human gut[122, 123], and soil[124] as potential sources - against the sequenced samples (as the sink). We mapped our reads against the greengenes database and imported them into QIIME for Sourcetracker analysis. Twenty calculus samples had at least 2500 reads that successfully mapped against the greengenes database, which is restricted to 16S rRNA genes, which were a minority of our metagenomics reads. The majority of our 16S sequences were of unknown origin (39-81%). Of the fraction of reads that were assigned by Sourcetracker, 12 samples resembled species from subgingival plaque, 5 were predominantly comprised of species resembling urban gut microbiota, and 3 had significant skin contributions as shown in Figure 1.

The three samples with high skin contributions in Sourcetracker were 81-100% Proteobacteria 16S genes, including several modern nosocomial pathogens. These three samples also grouped closely with the blanks, likely also due to the high proteobacterial content. Seventy-five percent of the 16S reads in one of these three samples, AMNH98173_Q01, mapped to an unclassified strain of *Acinetobacter baumannii*. In another of the samples identified by Sourcetracker to contain skin flora (KOD364459), the 16S reads mapped to *Stenotrophomonas maltophila* (44%), *Cronobacter sakazakii* (20%), and *Escherichia coli* (11%). The third sample, HUNT227684, had 16S reads that mapped to *Shewanella putrefaciens* (20%), an unclassified *Pseudomonas* strain (19%), *Aeromonas caviae* (15%), and *E. coli* (11%). Despite the high abundance of Proteobacteria in these three samples, all historical dental

calculus samples cluster more closely to each other, than to other modern microbial communities, such as gut, skin, and plaque microbiota.

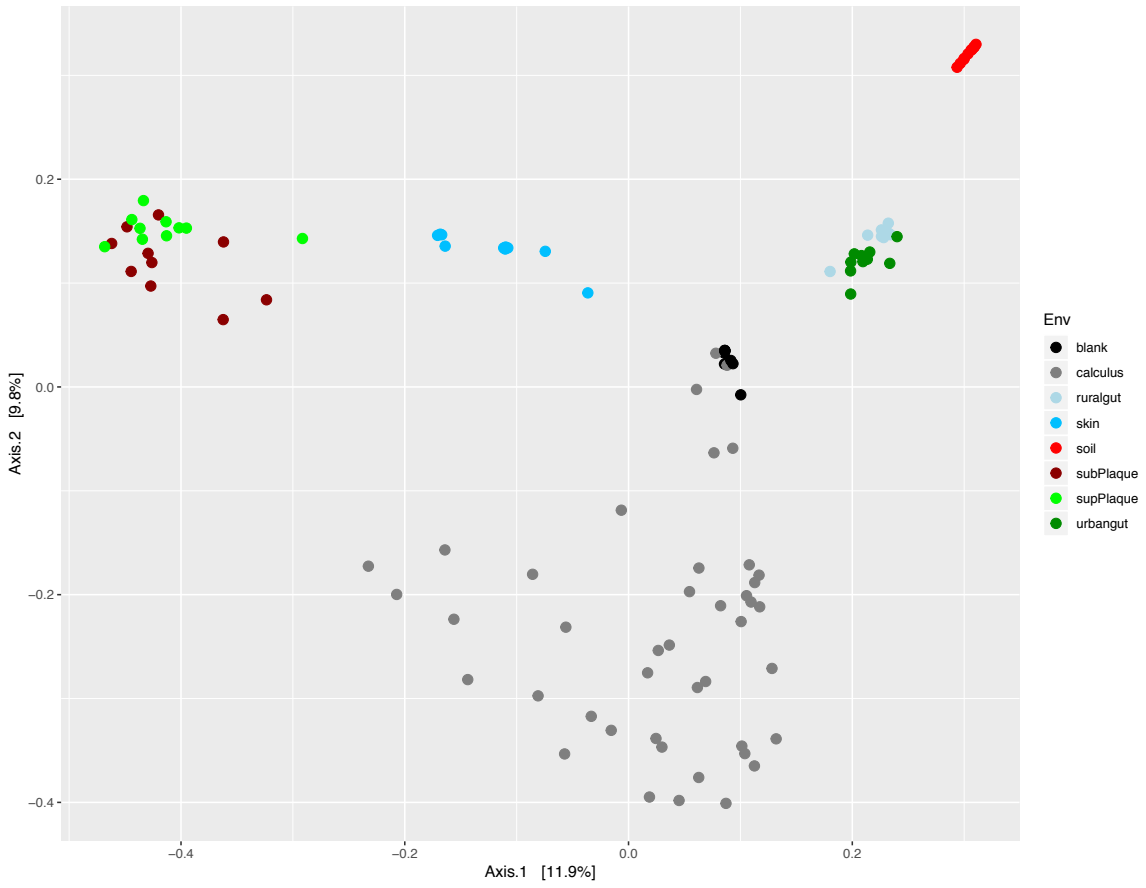


Figure 3.1: PCoA constructed using Bray-Curtis distances. Env (microbial environments) are human-associated microbiota, soil, and calculus samples from this study. Sourcetracker results are available on Supplementary Table 3.15

Read-based taxonomic analyses

A total of 675,937 reads were removed from samples following human contaminant analysis, leaving 225,096,484 for read- and assembly-based analysis. We observed a total of 1165 prokaryotic taxonomic categories and 1102 viral taxonomic categories. Metagenomic analysis was completed in both metaphlan2 (in Supplementary Information 3.8) and MEGAN (discussed below; metaphlan2 and MEGAN absolute abundance files outputs are available at Supplementary Tables 3.9 and 3.10 respectively) [125]. MEGAN showed prominent archaea within historical calculus samples. Bacteria accounted for 54-100% of assigned taxa while archaea represent 0-45% of the assigned taxa. We observed 25 phyla in historical calculus samples. Five phyla accounted for a majority of reads: Proteobacteria (average 37.42% per sample), Actinobacteria (average 22.51% per sample), Firmicutes (average 15.58% per sample), Euryarchaeota (average 14.22% per sample) and Bacteriodetes (average 5.64% per sample), as shown in Figure 2. Other phyla such as Chloroflexi, Spirochaetes, Synergistetes, Fusobacteria, and Candidatus Saccharibacteria, collectively account for on average 4.61% of the remaining reads.

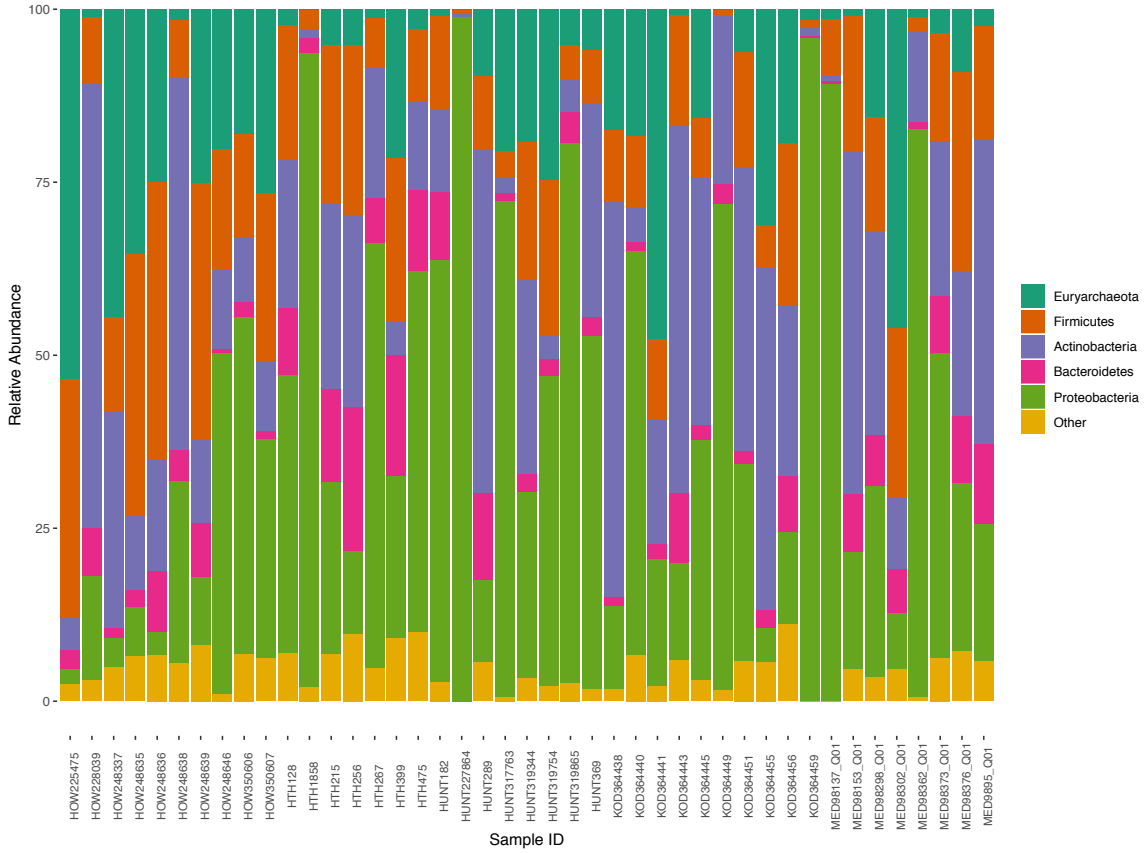


Figure 3.2: Stacked bar plot of dominant phyla across all historical calculus samples

We identified 578 species present with our historical calculus samples. Of which, *Methanobrevibacter oralis* dominated our historical calculus samples. *M. oralis* was represented on average 22.1% of samples and 0-74.35% of the reads from the 43 calculus samples analyzed were mapped to *M. oralis*. Other species represented in our samples include members of the “red complex” *Tannerella forsythia* (n=38/43; average 3.53% per sample) and *Treponema denticola* (n=28/43; average 0.48% per sample). There were also oral taxon representatives from Anaerolineaceae, Peptostreptococcaeceae, *Actinomyces*, *Olsenella*, and Lachnospi”race”ae.

In a total of 11 extraction, library, and QPCR negatives (“blanks”), we recovered 449 species level assignments. Bacteria dominate these control samples, being comprised of 22-91% Proteobacteria. In contrast, archaeal phyla such as Euryarchaeota and Thaumarchaeota account for less than one percent of recovered taxa (0.13-0.62% and 0.06-0.13%, respectively). Recovered genera include *Spingomonas*, *Pseudomonas*, *Methylotenera*, *Chrococcidiopis*, *Bulkholderia*, and *Parabulkholderia*. While there is overlap at the phylum level recovered in negatives and samples, no species were overrepresented in negative controls compared to positive samples and therefore no sequences were removed from the dataset. Additionally, *M. oralis*, the most dominant species among our historical calculus samples, was not detected in any of the negatives.

To define indicator species of historical calculus, our microbial communities were compared to modern plaque microbiota and 22 other ancient calculus samples (Supplementary Table 3.4). Of a total 545 species, 21 were associated with our historical calculus samples with high statistical support (p-values <0.005; shown in Figure 3). These taxa include *Methanobrevibacter*

species, *M. oralis* and *M. smithii*, *Desulfobulbus sp. ORNL*, *Actinomyces* sp, and *Olsenella* sp. Fifty-nine species were statistically associated with modern plaque samples [121] (Figure 3). These include *Streptococcus* sp, *Prevotella* sp., *Capnocytophaga* sp., and *Treponema* sp. In a comparison across both recent historical and more ancient microbial communities, 24 species were associated with our recent historical calculus samples (this study) with high statistical support, 44 species were associated with medieval European microbiota [27] and 2 species were associated with 18th century French calculus microbiota [108]. The two species distinguishing the 18th century French microbiota include *Streptococcus cristatus* and *Streptococcus sinenis*.

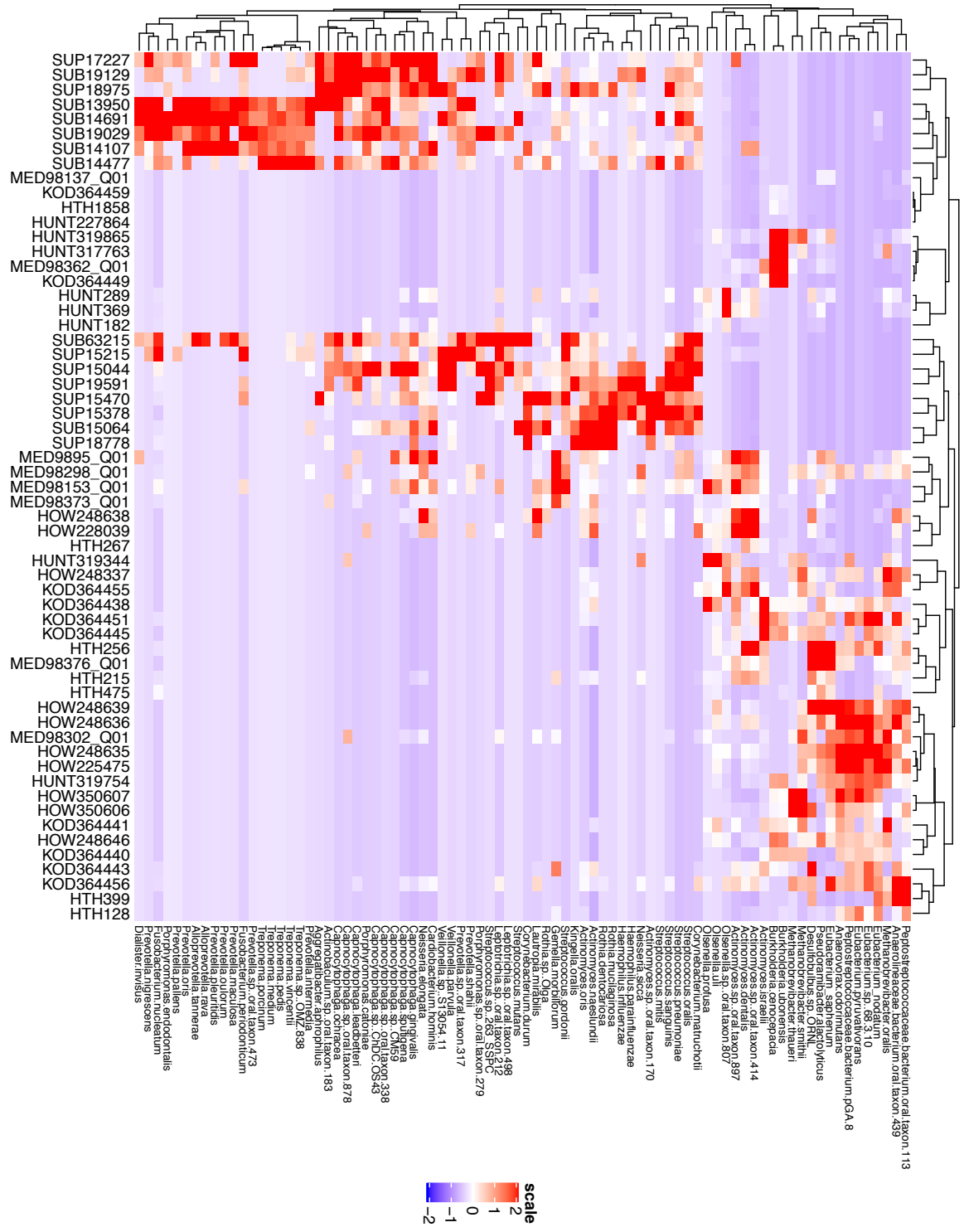


Figure 3.3: Heatmap displaying relative abundance of 80 indicator species across our 43 historical calculus samples and 16 modern supragingival (SUP) and subgingival (SUB) plaque HMP samples

Viral communities

Nearly all reads given viral assignments belonged to dsDNA viruses, including those that infect bacteria, archaea and eukaryotes. A comparison with viral reads identified in the negative controls (blanks) showed the same pattern as with the prokaryotic-mapping reads: none where overrepresented in negative controls than positive samples, and thus no adjustments were made to the datasets. Seven viral families accounted for 97-100% of our samples, in descending order: *Myoviridae*, *Siphoviridae*, *Phycodnaviridae*, *Mimiviridae*, *Podoviridae*, *Herpesviridae*, and *Marseilleviridae*. A total of 805 virus species (shown in Supplementary Table 3.9) were detected in dental calculus samples. *Bacillus G* virus (family *Myoviridae*) was the most prevalent species, present in 41/43 samples. Other prevalent species include *Geobacillus virus E3*, *Orpheovirus IHMI-LCC2*, *Cronobacter phage vB_CsaM_GAP32*, *Cafeteria roebergensis virus*, and *Clostridium phage phiCT453A*.

Few prokaryotic genera prevalent in historical samples had their phage counterparts prevalent in virome analysis, with the exception of *Actinomyces*, *Streptococcus*, and *Lactobacillus* phage. *Actinomyces phage* Acj61 which infects *Actinomyces massiliensis* was detected in our samples. *Streptococcus* phage infecting *Streptococcus equi*, *Streptococcus gordonii*, *Streptococcus oralis*, *Streptococcus pneumoniae*, and *Streptococcus suis* were also recovered. Although *Lactobacillus* was not resolved to a species level in our MEGAN analysis (which uses Lowest Common Ancestor to resolve blast hits to the lowest common ancestor), we observed 572 hits to *Lactobacillus* phage, including 44 hits to *Lactobacillus* prophage Lj771.

Viruses of Proteobacteria are also well-preserved within the historical calculus samples which had high potential skin contributions in sourcetracker analysis.

Acinetobacter phage Bphi-B1251 was recovered in sample AMNH98173_Q01. *Stenotrophomonas* phage Smp131 and *Stenotrophomonas* phage phiSHP2 were recovered in sample KOD364459. Additionally, *Aeromonas* virus phiO18P was recovered in HUNT227684.

Pathogen Reconstructions

Decontaminated reads were assembled into contigs in megahit for assembly-based analysis [109]. A total of 134,312 contigs from 41 samples (we did not assemble high quality contigs from sample HUNT319344, HTH1858 and HTH399 in our megahit analysis) were uploaded into the PATRIC database for metagenomic binning analysis [111]. Quality statistics for megahit-assembled contigs is available on Supplementary Table 3.12. We constructed 7 high quality bins of clonal pathogen populations, including three 100% whole genomes of *Methanobrevibacter oralis*, as shown in Table 3.4 (complete bin statistics is available on Supplementary Table 3.13). Three reconstructions of *M. oralis* genomes are shown in Figure 3.5.

Genome Name	SampleID	Overall alignment rate (%)	Coarse Consistency (%)	Fine Consistency (%)	Completeness (%)	Contamination (%)	Mean coverage
Acinetobacter baumannii clonal population	MED98137_Q01	35.33	98.2	92.8	95.3	7.2	7.84
Anaerovorax odorimutans clonal population	HOW248639	0.09	92.8	91.4	83	5.1	12.35
Anaerovorax odorimutans clonal population	HTH128	0.06	96.3	95.5	95.9	4.2	10.13
Denitrobacterium detoxificans clonal population	HOW248337	0.06	97.3	94.7	100	6.8	20.96
Methanobrevibacter oralis clonal population	HOW248337	17.71	99.1	96.5	100	7.8	93.99
Methanobrevibacter oralis clonal population	HOW225475	16.44	99.5	99.3	100	0	59.81
Methanobrevibacter oralis clonal population	HOW248635	13.55	99.4	98.7	100	1.6	69.01
Stenotrophomonas maltophilia clonal population	KOD364459	26.63	97.7	91.4	98.1	9.3	16.94

Table 3.4: Metagenomic bins constructed in PATRIC. Summary statistics include coarse consistency (%; percent of predictable roles whose presence/absence matches the computations of the consistency tool), fine consistency (%; present of predictable roles whose number of occurrence matches the computations of the consistency tool), completeness (%), contamination (%; extra material in the bin), and mean coverage (average coverage for contigs in the bin) provided by PATRIC. Contigs in metagenomic bins were mapped back to the PATRIC reference and overall alignment rate is reported in the table above. Complete bin statistics are available in Supplementary Table 3.13. More details available at: https://docs.patricbrc.org//tutorial/metagenomic_binning/metagenomic_output.html

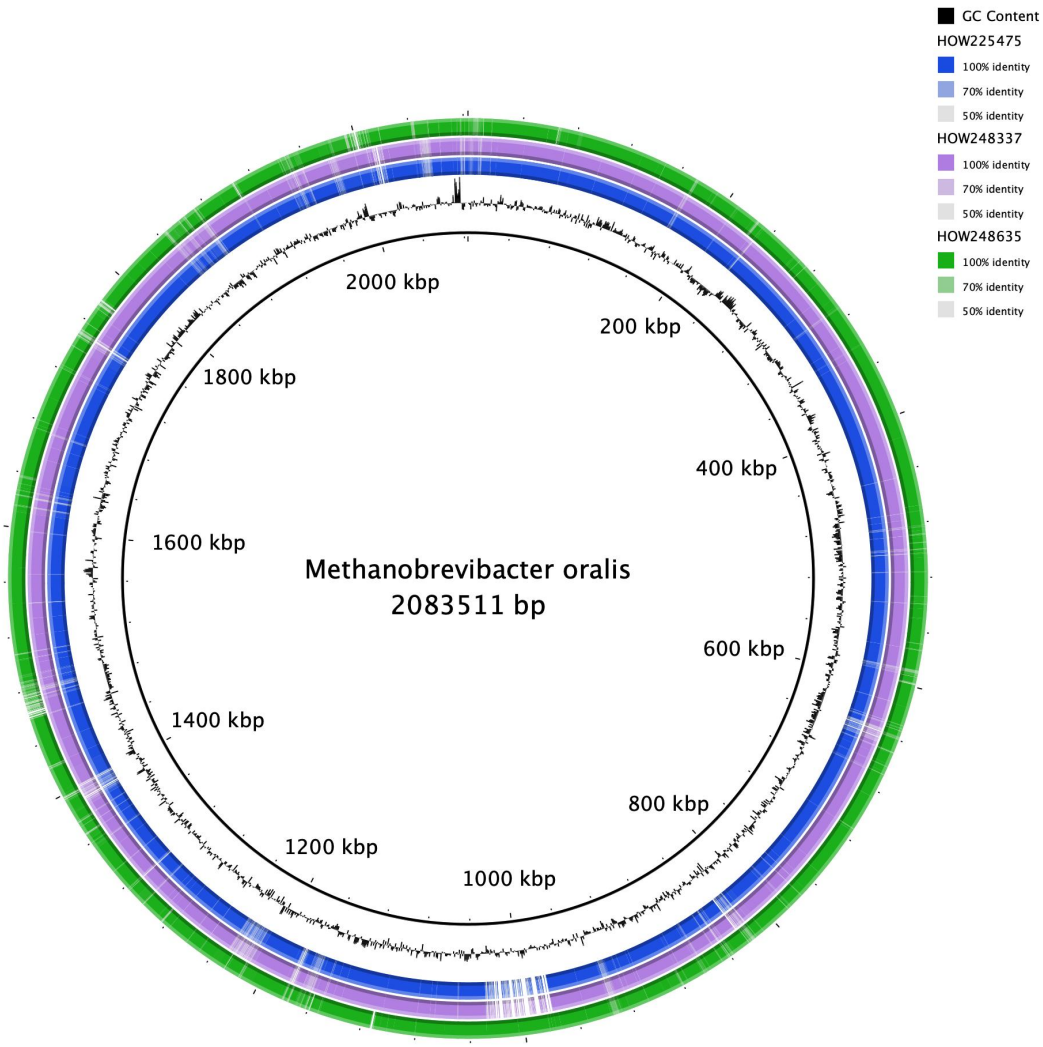


Figure 3.5: Circular genomes constructed in BRIG of three reconstruction of *Methanobrevibacter oralis*. The reference strain is in the center, the HOW225475 strain in blue, the HOW248337 in purple, and HOW248635 strain shown in green. Note, all reconstructed strains fell within the same pangenome cluster.

The GC content for reconstructed *M. oralis* ranged from 28.32-28.71%. Each genome was comprised of 31-34 tRNAs and encode for 2099-2550 proteins. CRISPR arrays were detected on two out of three complete reconstructed genomes. Additionally, specialty genes for antibiotic resistance and transporters were present in reconstructed strains. All reference strains had genes encoding for isoleucyl-tRNA synthease (EC 6.1.1.5) in the mupirocin class. D-alanine-D-alanine ligase (EC 6.3.2.4) in the D-cycloserine class was also present on a reconstructed strain.

Prophages were also detected in reconstructed *M. oralis* strains. Predicted viral insertions in *M. oralis* strain from HOW225745 on position 1537673-1569719 is homologous to four genes present in archaeal prophage *Methanothermaobacter wolfeii PsiM100* (NC_002628). A 9.2kb region within the reconstructed *M. oralis* strain from HOW248337 also contains an incomplete prophage with homology to *Spingomonas phage PAU* (NC_019521). Three incomplete phage insertions were found within the reconstructed strain from HOW248635 which were 7.1kb, 8.3kb, and 8.5kb in length respectively. *Aeromonas phage 25* (NC_008208), *Prochlorococcus phage P-TIM68* (NC_028955.1), and *Bacillus phage G* (NC_023719) are the taxa with the highest number of proteins similar to region for each of three regions respectively.

Upon comparing the reference *Methanobrevibacter oralis* strain to our three reconstructions, we observed many structural variants. In the HOW225475 *M. oralis* strain, there were 2238 snps, 132 indels, and 309 complex variants. A majority of complex mutations and snps in HOW225475 were in DUF (domain of unknown function) proteins. CRISPR-associated, transposases, restriction endonucleases, and multidrug transporter proteins are among the mutated

genes in this strain. In the HOW248337 strain, there were 3401 snps, 222 indels, and 260 complex mutations. In addition to the aforementioned genes, mutations with stage II sporulation protein M-like proteins, integrase, co-/chaperones GrpE, DnaK, DnaJ, and arsenic resistance proteins are among the complex mutations and snps. In the HOW248635 strain, we observed 3958 snps, 199 indels, and 613 complex mutations. Nitrogen reductase, MATE family efflux transporters, nucleotide sugar epimerase, and prepilin peptidase are among the complex mutations and snps.

Pangenome Analysis

Further analysis of the pangenome of *M. oralis* resulted in the more comprehensive analysis of available data. Seventeen samples (median coverage 2.86-146.13), which also included HOW225475, HOW248337, and HOW248635, were analyzed along with available reference sequence data on GenBank. As of 3/2019, only 3 genomes were available from comparative analysis sampled from the human mouth, stool, and colostrum. We recovered 1792-1934 gene families in historic era samples and an average of 1975 across 3 reference strains.

We observed 5 distinct clusters based on presence-absence of pan-gene families shown in Figure 3.4. The first two clusters were solely comprised of modern reference data, while the other three clusters were solely comprised of historical calculus samples. The first cluster was comprised of gene families responsible for transcription regulation, nucleoside-diphosphate sugar epimerase, and ribonuclease MRP protein subunit RMP1. Two gene families present in the second cluster, but absent in all others include DDE transposases and a conjugal transfer protein. The third cluster also has two large segments of

missing gene families present in all other clusters. While a majority of which are hypothetical proteins, the remaining gene families encode integrases, endonucleases, and a phage portal protein. Upon further analysis of gene families with known protein function, we recovered 100% homologous sequence to M. oralis when nucleotide sequences of site-specific integrases, one HNH endonucleases, and one phage portal protein. One HNH endonuclease (gg001816) has 99.51% percent identity to *M. oralis*.

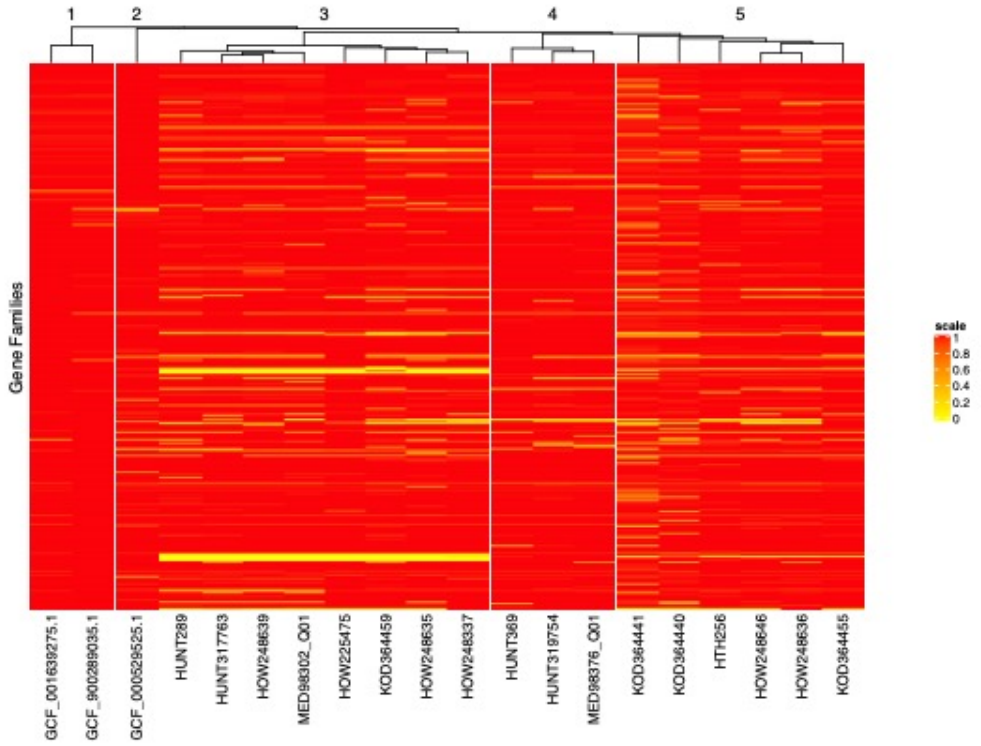


Figure 3.6: Heatmap of pan-gene families across 17 samples. Here, we show 5 distinct clusters of *M. oralis* using pairwise distances from a gene family presence-absence matrix.

Co-occurrence Analysis

A total of 1,307 positive and 362 negative relationships were found in total. Of these, *M. oralis* was found in 13 associations, 10 of which cooccurred with *M. oralis* in over 30 samples (probabilities >0.5). These associations include taxa associated with historical calculus samples in indicator species analysis: *Olsenella* sp., *Actinomyces dentalis*, and *Desulfobublus sp. ORNL*. Additionally, known periodontal pathogens *Tannerella forsythia* (prob=0.84) and *Treponema denticola* (prob=0.619) were observed to cooccur with *M. oralis* in 37 and 28 samples respectively.

Discussion

DNA quality across sample groups

In our study, we provide the first multi-museum comparison of contemporary geographically and racially diverse Americans that died at the turn of the century. Poorly preserved DNA is an artifact of specimen acquisition, treatment, and storage which differ across museums under study [39]. Skeletal collections curated by Aleš Hrdlička at the NMNH were defleshed by boiling in steam-piped vats and left to dry to dissecting tables [40] while anthropologists at the CMNH first embalmed the remains, treated them with chemicals, and them boiled them [126] . Although museum collections have been reported to have low amounts of DNA preserved within calculus, 43/44 samples yielded study had sufficient unique reads (>500,000) for metagenomics analysis [39].

The Historical Calculus Microbiome

When compared to other microbial habitats, calculus samples formed a distinct cluster, away from soil, skin, gut, and plaque sources (Figure 3.1). This

trend also appeared in a recent paper about the dental calculus microbiota of Italians who died between the 11th and 19th centuries [23]. Additionally, historical calculus samples are widely distributed, showing dissimilarity between samples, while modern plaque samples are more tightly clustered. This pattern illustrated was also described in periodontal metagenomes, where diseased samples cluster more tightly while healthy samples are more dissimilar [127]. This supports previous hypotheses of pre-modern microbiota being more diverse than modern day plaque microbiota [22].

While numerous factors may drive the dissimilarity shown, it is not demographic factors such as “race” or place of geographic origin (Supplementary Figures 3.14-17). Dental paleopathology, number of caries per tooth, antemortem tooth loss, and dental wear, of contemporaneous Americans also do not reflect patterns associated with “race”. Dental health and calculus microbiome data also suggest that dental health of historic-era Americans were similar to each other and patterns associated with dental health disparities may be a recent phenomenon.

Historical calculus samples were comprised of microbiota most similar to that found in modern subgingival plaque with a large proportion of unknown microbial contribution (Supplementary Table 3.19). Unknown microbial contributions to samples included an average 25% of unclassified 16s and archaeal genera *Methanobrevibacter*. The remaining bacteria contributing to the unknown signal in our samples included *Actinomyces*, *Desulfobulblus*, *Eubacterium* and *Pseudoramibacter*, genera of historical calculus indicator species.

"The Archaea Effect" [25]

Methanobrevibacter oralis was first described in 1994, isolated from human subgingival plaque [128]. Subsequently, *M. oralis* was isolated in human colostrum and stool [129]. *M. oralis* was nearly ubiquitous in our calculus samples and among the 22 indicator species associated with historical calculus. This trend has been reported in historical oral microbiota, accounting for up to 60% of reads recovered from 17th -19th century from Dutch dental calculus microbiota [25]. However in healthy ancient dental calculus microbiota, *M. oralis* is notably absent [23]. *M. oralis* is also absent in healthy modern supragingival and subgingival plaque. Methanogens, including *M. oralis*, may play a role in the etiology of periodontal disease as it is often isolated in periodontal sites and absent in healthy sites [130]. The production of methane in the human periodontal pocket may increase total microbial activity by consuming the metabolic byproducts of anaerobic bacteria thereby increasing local tissue damage in the periodontal pocket [131].

M. oralis is considered a "co-pathogen", found in association with other disease-causing microbes [132]. In 38 out of the 43 samples that underwent co-occurrence analysis, *M. oralis* was found in strong association with *Tannerella forsythia*, *Anaerolineaceae oral taxon 439*, and *Desulfobublus sp. ORNL T. forsythia* is a member of the "red complex", a consortium of periodontal pathogens [133]. *T. forsythia* has been characterized in other calculus microbiota and a complete genome was reconstructed in a study of periodontal calculus from medieval Germany [22, 27]. There are currently only 2 nucleotide sequences from *Anaerolineaceae oral taxon 439* that were directly submitted to NCBI. Anaerolineaceae are Gram-negative, non-spore-forming, and nonmotile members of Chloroflexi [134]. It is considered an opportunistic pathogen, found

within gout fecal microbiome and oral squamous cell sarcoma [135, 136]. They are also methanogens and compete with *M. oralis* for hydrogen sources within the oral cavity. *Desulfobulblus sp. ORNL*, renamed *Desulfobulbus oralis*, is a sulfur-reducing deltaproteobacterium and a potential syntrophic partner for *M. oralis* [137].

The differential detection of archaea in modern and ancient microbiome data may be due to methodological issues, such as DNA extraction protocols that do not lyse archaeal cell walls [138]. The failure of previous widely used protocols led to the underestimated to other members of *Methanobrevibacter* in stool, *M. smithii* and *M. stadtmanae*. Archaeal specific primers may also aid in capturing the archaeal diversity [139]. Previous research has also shown that taphonomy may impact the recovery of *Methanobrevibacter* from dental calculus using 16S amplicon sequencing [25]. *Methanobrevibacter* has the shortest V3 region of oral taxon; it is 17 base pairs shorter than the shortest bacterial sequence [25]. Shorter DNA fragments are easy to recover in ancient DNA datasets, which may lead to the overrepresentation of Methanobrevibacter compared to bacterial counterparts [25]. However, we found that 16S analysis conducted in QIIME lead to the underrepresention of *Methanobrevibacter* compared to alignments in bowtie2, metaphlan2, and MEGAN analysis (Supplementary Figure 3.19). Therefore, our results show that *Methanobrevibacter* is a substantial part of recent historical oral microbiomes and do not show the previously observed taphonomic bias for archaea. Secondly, ancient DNA degradation have been associated with GC content and DNA fragment length changes [140]. Shorter DNA fragments tend to have higher GC content, which again may lead to the absence or underestimation in ancient

dental calculus. The GC content of *Methanobrevibacter oralis* is 27.70-28.00% in reference strains and 28.32-28.71% in reconstructed strains.

Thirdly, both modern and ancient microbiome research suffers from sampling bias towards people of European ancestry, missing population-specific differences in archaeal carriage [141-143]. *M. oralis* is known to be differentially abundant across different ethnicities [130, 144-146].

The potential overrepresentation of Archaea within historical calculus samples may also be attributed to the loss of syntrophic partners for *M. oralis*. Both *Anaerolinaeceae oral taxon 439* and *Desulfobulbus oralis* are indicator species, associated with historical calculus microbial communities with high statistical support. *Anaerolineaceae oral taxon 439* is present on average of 2.71% and *Desulfobulbus oralis* is present on average 3.56% of historical calculus samples. Both of these taxa are absent in modern supragingival and subgingival plaque metagenomes recovered from the HMP. Additionally, there is evidence that the prevalence of methanogens have increased over time [147]. For all of these reasons, it is likely that we missing the breadth of archaeal diversity in the human oral microbiome.

Our pangenome analysis revealed the rich diversity of *M. oralis* and reconstructed strains doubled the amount of *M. oralis* genomes available to the public on GenBank. Two of the five clusters were only comprised of reference strains with modern subgingival plaque and colostrum apart from stool. All historical samples were in the remaining three clusters and did not cluster by demographic data, such as collection, geography, or “race”. Our three reconstructions fell within a single cluster, missing gene families responsible for horizontal gene transfer. The oral cavity has a greater diversity of mobile genetic elements than the gut [148] and thus clusters may be defined by gene

transfer experienced by a single individual based on their unique microbiome and not shared demographic traits.

The Historical Oral Virome

The oral virome is dominated by bacteriophage, though eukaryotic viruses belonging to *Herpesviridae*, *Papillomaviridae*, and *Anelloviridae* may be present in asymptomatic healthy individuals [149-152]. Virome research on modern day peoples find the oral viral populations are not transient, but are individual-specific and remain stable and persist over time [150, 152, 153]. In healthy mouths, the virome is dominated by family *Siphoviridae* in both health and disease. A majority of members of *Siphoviridae* have a lysogenic lifestyle [153]. In periodontal disease, *Myoviridae* dominates subgingival plaque [153]. *Myoviridae* is comprised of lytic phages, which suggests these viruses might play a more active role in maintaining bacterial populations in periodontal pocket [153]. *Bacillus G* virus, in the family *Myoviridae*, is the most prominent virus in our study and present in 41/43 samples. We also found an integrated *Bacillus G* phage in one of our *Methanobrevibacter oralis* reconstructions.

Similar to published datasets, dominant viral families in this study included *Siphoviridae*, *Myoviridae*, and *Podoviridae* [150-153]. *Phycodnaviridae* are also among the most dominant viral families, present on average 9.01% of each sample. *Phycodnaviridae* is a family of viruses that infect marine and freshwater algae. Among the most abundant reads in this family are *Only Syngen Nebraska virus 5*, which infects green alga *Chlorella variabilis*. Other constituents of *Phycodnaviridae* detected in our samples were *Chlorovirus*, *Coccolithovirus*, and *Prasinovirus*. These genera were not detected in our negatives and they are not known residents of oral microbiota. We believe

Phycodnaviridae are transient viruses, encapsulated within dental calculus as a result of ingestion [154]. The consumption of seaweed is well-documented in 19th and early 20th century cookbooks [155-157]. Seaweed was a food of last resort for poorer classes when crops failed [158]. In the grocer's encyclopedia, written in 1911, there are several entries featuring seaweed: carragheen, kanten, dulse, kelp, and laver [159]. It is high in protein and is boiled and served with butter, vinegar, and fried bacon or fish or served with cold meats [159]. It is also used to clarify beer and as a demulcent when mixed with water, milk, sugar and/or lemon juice [159]. Seaweed has also been used as fertilizer and animal fodder [154].

We can also not exclude the poor quality of water and inadequate filtration systems at the turn of the century. Water filtration dates to the Bronze Age, but was not successfully implemented in the USA until 1832 [160]. Early water systems were ineffective at maintaining the separation of water and waste [161]. To improve quality and taste and disinfect water systems, chlorination was implemented on a large scale at the Booton Reservoir in Jersey City in 1908 and implemented by other major cities within the next decade [161]. However, for Americans that lived in rural or low population areas, like Kodiak Island, Alaska, for example, community health infrastructure was nonexistent [162]. The time period under study predates the death of a majority of individuals in our study so it unlikely to be the source of algae-infecting viruses.

Thirty-three bacterial and archaeal species were strongly associated with historical calculus communities compared to ancient and modern oral microbiota. Interestingly, these did not include members of *Streptococcus* and *Actinomyces* prominent constituents of modern oral microbiota. However, *Streptococcus* and *Actinomyces* phage were present in historical calculus

samples. Additionally *Mycobacterium* accounted for less than 1% of microbiota recovered from individuals and we found 35 species of *Mycobacterium* phage that infect *M. tuberculosis*, *M. avium*, and *Mycolicibacterium smegamatis*.

Historic-Era Pathogens

Dental calculus metagenomes can provide insight into systemic pathogens, such as *Acinetobacter baumannii* and *Stenotrophomonas maltophila*, which we reconstructed in our study. The presence of modern nosocomial pathogens within dental calculus has been described in previous studies[23]. *A. baumannii* is a current nosocomial pathogen and the cause of endemic and epidemic outbreaks in hospitals [163, 164]. Mobile genetic elements carrying antibiotic resistance are common in modern isolates of this pathogen and may have been recently acquired from *Escherichia*, *Pseudomonas*, and *Salmonella* bacteria [165, 166]. Members of these genera were not recovered from this individual. Our reconstructed *A. baumannii* strain from AMNH98137_Q01 an individual who died between 1945-1946 harbored a gene for resistance to carbapenem, a class C beta-lactam introduced in 1985 [163]. It is believed the introduction and overuse of carbapenems led to the development of resistance in *A. baumannii* [165], however all modern isolates of *A. baumannii* encode a homologous class C beta lactamase (ADC, *Acinetobacter*-derived cephalosporinase) on their chromosome [167]. The presence of a carbapenem resistant phenotype of *A. baumannii* in an individual who died in the pre-antibiotic era suggests this phenotype predates the antibiotic revolution, and confirms that the ubiquity of the ADC gene in *A. baumannii* is due to its long presence in the genome rather than a recent selection event.

A single individual, KOD364459, experienced multiple infections during the time the calculus we sampled from was formed. The genome of *Stenotrophomonas maltophila*, another modern nosocomial pathogen, was reconstructed from this individual. The reconstructed strain shared gene families with strains that isolated from the human brain and the oropharynx, although *S. maltophila* can be found in a variety of environmental sources as well [168]. Additionally, *E. coli* and *C. sakaskii* were also present in high abundance.

Conclusion

In conclusion, historical microbiota appears to have a different microbial composition than modern day plaque communities. While there is overlap in their microbial ecology, results show an abundance of archaeon *Methanobrevibacter oralis* in historic-era calculus. *M. oralis* co-occurs with other microbes that are indicator species, such as *Tannerella forsythia*, *Anaerolineacea*e oral taxon 439, and *Desulfobublus* sp. ORNL, strongly associated with historical calculus samples. Additionally, there are differences in beta diversity that suggest that historic-era calculus was more diverse than modern day plaque microbiota. While *M. oralis* dominant-microbiota appeared more diverse, perturbed microbiota was dominant in Proteobacteria pathogens, such as *Acinetobacter baumannii* and *Stenotrophomonas maltophila*. In addition to microbiota, we characterized prevalent metabolic pathways and genes that encode virulence, antibiotic, and metal resistance in our samples. While this study provides support to the belief that modern day microbiota is less diverse and in a diseased state, more work is required to determine the applicability of our results across Americans of all socioeconomic status at the time.

CHAPTER 4: EFFICACY OF DENTAL CALCULUS IN RECONSTRUCTING ORAL MICROBIOTA

Abstract

Dental calculus (tartar) is calcified dental plaque that can be found in living human populations as well as fossilized apes and hominins. Like plaque, calculus contains a complex microbiome that is preserved within its matrix and samples as old as 48,000 years old have been analyzed to reveal the oral microbiota of our closest extinct relative, *Homo neanderthalensis*. However, the relationship between dental plaque microbiomes, which are well-studied, and dental calculus, which has been exclusively studied in historical samples, has never been established. Here we conduct the first comparison of plaque and calculus in ten modern, healthy humans. Whole genome metagenomics revealed very similar species present in both calculus and plaque (from the same individual), though without a perfect overlap between them (92.42%-94.1%). Supragingival plaque samples from the human microbiome project (HMP) were most similar to dental calculus samples when compared to other HMP sites, and, unsurprisingly, plaque from the same individual was even more similar to their own calculus. Our results show that modern calculus is as diverse as the HMP microbiomes, but with age the ability to detect these diverse species in calculus may decline. Our results assert the dental calculus has a distinct microbial community with strong similarities to dental plaque and justifies the use of dental calculus in ancient and historical studies of the human microbiome.

Introduction

Investigations of ancient and historical human oral microbiota have been based almost entirely on dental calculus. The oldest deposits of dental calculus date to Miocene apes (12.5-8.5MYA) [169] and fossil hominins such as *Homo naledi* and *Homo neanderthalensis* [170, 171]. Dental calculus is ubiquitous in historical human populations and over 80% of adults today have calculus [16]. While ancient microbiome research is infancy, we are beginning to understand the effects of diet, lifestyle, and culture on microbial population and structure [22, 82]. However, without modern dental calculus microbiome data, comparisons to modern dental plaque are problematic.

Dental plaque is a polymicrobial biofilm [16]. The development of dental plaque on dentition occurs in several stages [16]. First, about 2 hours after brushing, the acquired enamel pellicle is formed and is comprised of salivary glycoproteins, phosphoproteins and lipids, and host defense components [16, 172]. Salivary proteins undergo conformational changes, exposing receptors that facilitate bacterial attachment [16, 172]. Second, initial microbial colonizers such as *Streptococcus sanguinis*, *S. oralis*, and *S. mitis* bind the receptors on the acquired pellicle [16, 172]. Third, other early colonizers, such as *Actinomyces* sp., and late colonizers, such as *Eubacterium* sp. and *Prevotella* sp., bind to receptors of early colonizers [16]. Microorganisms, such as *Fusobacterium nucleatum*, serves as a “bridge”, co-aggregating both early and late colonizers [16]. Finally, the biofilm continued to multiply, mature, and diversify, accumulating up to 200 species in any one individual [16, 172, 173].

Dental calculus, on the other hand, is calcified dental plaque. Dental plaque periodically mineralizes from calcium phosphate ions in saliva and

gingival crevicular fluid [174, 175]. If not removed, the calculus continues to build to form a “cement-like substrate” similar to that found in bone and dentine [175]. High calculus formers tend to have high concentrations of calcium ions in their saliva and blood [16, 174]. Salivary flow rate, fluid consumption, and diet may also attribute to the formation of dental calculus while dental hygiene practices can help abate it [16]. While dental plaque is comprised of living and multiplying bacteria, dental calculus is comprised of dead inactive bacteria, foodstuffs, and endogenous DNA [175].

In this study, we sought to characterize the dental plaque and dental calculus microbiota of the same healthy individuals. Given the likely overlap in structure of dental plaque and calculus communities, we expect them to be comprised of the same bacteria at the same relative abundance. When compared to other oral microbial communities, we hypothesize that dental calculus will be more similar to dental plaque than other oral microhabitats. We believe this study will have relevance for evolutionary and medical analysis of oral microbial communities and implications for the detection and prevention of oral and systemic diseases.

Materials and Methods

Subject Recruitment and Selection

Ten human subjects were recruited at the Clinical Research Center at Rutgers University, Newark, NJ according to approval IRB protocol (Pro20170001450). The clinical examiner obtained subject consent and performed all clinical procedures. Subjects meeting the inclusion criteria and not having the exclusion criteria were asked to review and sign the informed consent. We sought adults aged 20-64 in good dental health. We excluded

individuals based on the incidence of chronic dry mouth, the occurrence of periodontal pockets equal to or greater than 4 mm, the appearance of cancerous lesions, oral candidiasis, halitosis, and individuals with more than 8 missing teeth. Subjects who signed our informed consent form and met inclusion criteria were examined for the presence of calculus on their upper and lower anterior teeth. Subjects enrolled in our study were asked refrain from tooth brushing and flossing 12 hours before the scheduled sample collection visit.

Age, sex, and self-identified “race” were recorded at the time of the subject’s sample collection appointment. The subject’s ages ranged from 20-42. There were 4 males and 6 females including in our sample. Subject 5 did not return for sampling and so an 11th person was recruited in their place. A calculus sample measuring approximately 1.5 mm at its longest dimension was removed from one of the supragingival surface of the anterior teeth with a sterile curette (Columbia 4R/4L, Hu-Friedy Instrument Company, Chicago, IL) and placed in a sterile 2 ml microfuge tube with 1 ml of 0.5M EDTA. Dental plaque was also collected from the labial surface of anterior teeth using a sterile curette (Columbia 4R/4L, Hu-Friedy Instrument Company, Chicago, IL) which was then placed in a sterile 2ml microfuge tube with 1 ml of 0.5M EDTA.

Sample Decalcification and DNA Extraction

Dental calculus and plaque samples were washed with 1 ml of 0.5M EDTA, placed on a nutator for 15 minutes, and centrifuged at 13krpm for 3 minutes. The supernatant was then transferred to a new tube and 1 ml of 0.5M EDTA was added to the pellet and placed on the nutator for 24 hours. After 24 hours, 100ul of Qiagen Proteinase K was added to each sample. Samples were placed back on the nutator for further decalcification for 48 hours. We used the

Qiagen DNeasy Powersoil kit with modifications to process lysed DNA (modifications discussed in Supplementary Information 4.1). Shotgun libraries were prepared by the Rutgers Genome Cooperative (http://dblab.rutgers.edu/genome_cooperative/). gDNA samples (measured by Qubit (dsDNA-Broad Range) were diluted to 0.2 ng/ul, and libraries were prepared using the Illumina Nextera XT DNA Library Prep Kit, following the manufacturer's protocol. For libraries that measured "too low" on Qubit, 5 ul of undiluted sample were used for library prep. Library quality was assessed using Qubit (dsDNA-Broad Range) and Bioanalyzer 2100 High Sensitivity DNA Chip. The only samples that did not yield any final library fragments were the negative controls, EN1 and EN2, so they were not sequenced. Libraries were pooled in an equimolar fashion at a concentration of 15 nM (10 ng/ul for 1 kb library fragments). The pooled libraries were sent to GENEWIZ for sequencing on one lane of an Illumina HiSeq 4000, with a 2x250bp configuration. A detailed flow chart of all wet laboratory work is available in Supplementary Figure 4.2.

Bioinformatics Analysis

First, we checked sequencing quality of all samples using FASTQC version 0.11.8. Sequencing adapters were removed, reads were merged, and quality filtered in AdapterRemoval version 2.1.7 [97]. Human-associated DNA was removed using in kneaddata (https://bitbucket.org/biobakery/kneaddata/wiki/Home). Metagenomes were analyzed in two different programs, MEGAN [105] and metaphlan2[98]. Decontaminated reads were first analyzed in metaphlan2 using default parameters. Then, metagenomic sequences were aligned using DIAMOND to bacterial and archaeal reference protein sequences and all viral proteins

available on GenBank (downloaded 3/2019, 70,800,065 amino acid sequences) [104]. Both decontaminated sequences and resulting diamond files were used to create RMA files for analysis in MEGAN v.2.15.62 [105]. For comparative purposes, historical dental calculus (from Chapter 3) metagenomic sequences (processed in the same manner) were run in tandem. Additional modern samples from human oral, nasal, throat samples were also downloaded from the HMP browser (https://www.hmpdacc.org/resources/data_browser.php; list shown in Supplementary Table 4.3) [121]. Periodontal plaque metagenomes were also downloaded from Genbank (list of sequences used in analysis in available in Supplementary Table ?.?).

Statistical Analysis

The core microbiome for all samples in our analysis was calculated at 50%, 75%, and 100% sample threshold in MEGAN [105]. Differentially abundant biomarkers were analyzed using LEfSe [176] using the LDA of 2 as a cutoff. Species level taxonomy tables were converted into presence-absence tables and analyzed in R package indicspecies for indicator species analyses [118]. Species with p-values <0.005 were included in downstream analysis. Sample metagenomes were downloaded as BIOM files from MEGAN and imported into QIIME for source-tracking analysis was conducted using sourcetracker v1.0.1 [102, 103]. Sourcetracker analysis was conducted using the ‘sourcetracker_for_qiime.r’ script from v.1.0.1 release [103]. We ran two analysis using Sourcetracker to determine the composition of calculus samples 1) using the same individual’s plaque and samples from all HMP oral microhabitats and 2) using HMP supragingival and subgingival plaque samples. In both analysis, anterior nares, buccal mucosa, keratinized gingiva (gingiva

adjacent to the teeth), palatine tonsils, saliva, throat, and tongue dorsum samples were analyzed as sources. All data was rarefied to 10,000 reads. Prior to diversity analysis in RStudio, taxa were filtered so that only those that had a relative abundance with a mean greater than 10^{-5} (to remove low abundant taxa) were kept in alpha diversity analysis in phyloseq v1.19.11 [177, 178].

A detailed flow chart displaying the bioinformatic workflow is shown in Supplementary Figure 4.4.

Results

Sequencing Quality

A total of 10 dental calculus and 10 dental plaque samples generated 88,789,783 reads that were on average 288bps in length. Following adapter removal, quality filtering, trimming, and merging of paired ends, we had 55,776,193 reads for decontamination analysis. Prior to metagenomic analysis, human contaminant reads were removed using kneaddata. Human contaminants accounted for .009-81% (min 25,877 reads-max 3,121,252 reads) of merged, trimmed reads. Human contamination varied between dental calculus and dental plaque, with an average of 25.31% (0.9-62%) and 39.2% (11-81%) human contamination respectively. The remaining 35,665,629 reads were used for downstream analysis.

Taxonomic Composition

Both dental plaque and calculus metagenomes were dominated by four phyla in MEGAN: Bacteriodetes, Actinobacteria, Proteobacteria, and Firmicutes (as shown in Figures 4.1 and 4.2). Bacteriodetes was present on average 17.42% of calculus and 11.89% of dental plaque. Actinobacteria accounted for

on average 45.19% of dental calculus and 42.92% of dental plaque. Proteobacteria was present on average 21.44% of dental calculus samples and on average in 11.04% of dental plaque samples. Firmicutes accounted for 12.88% of dental calculus and 31.78% of dental plaque samples as shown in Figure 1. Candidatus Saccharibacteria, Fusobacteria, Spirochaetes, Euryarchaeota, Synergistes, and Chloroflexi accounted for on average 0.51% and 0.37% of dental calculus and dental plaque respectively.

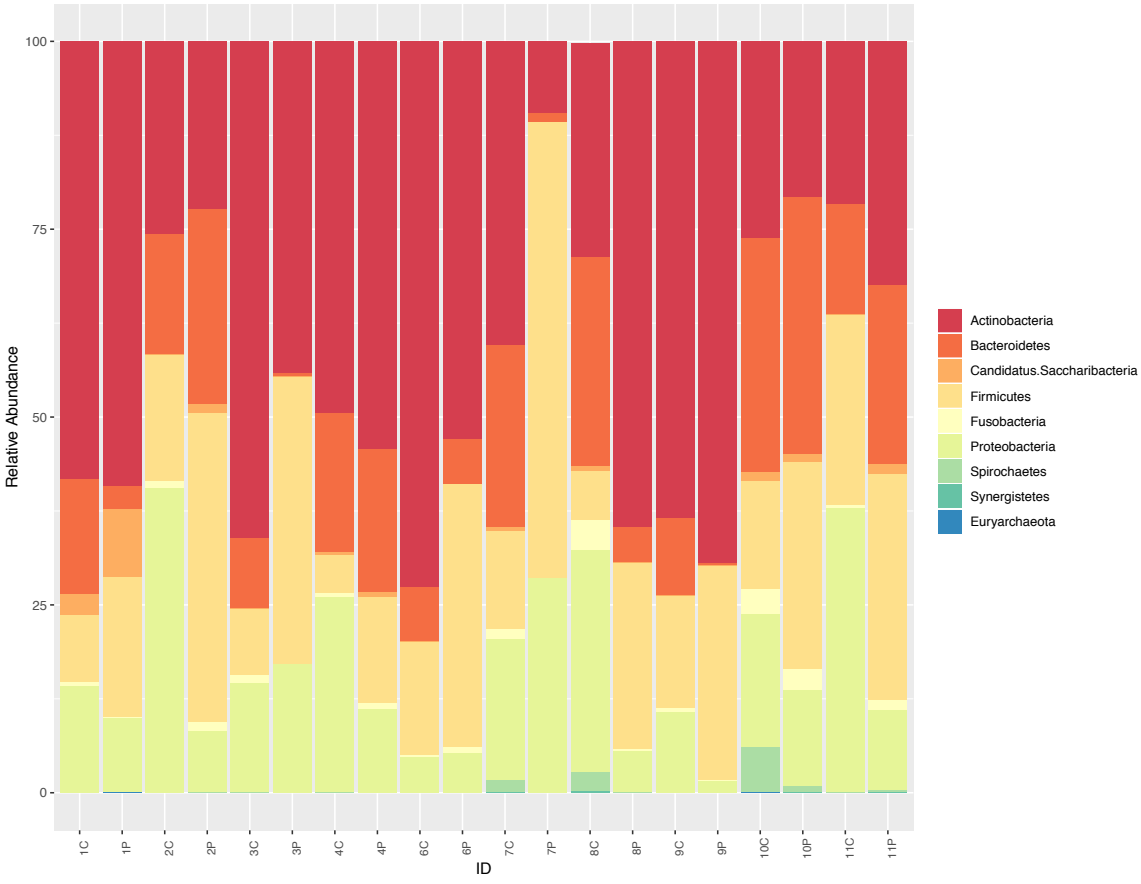


Figure 4.1: Dominant phyla across all samples as assigned by MEGAN. Individuals are numbered 1-11 (there is no sample 5), marked C for calculus and P for plaque.

All samples were also re-analyzed in metaphlan2. Dominant phyla are similar to those found by MEGAN: Bacteriodetes, Actinobacteria, Proteobacteria, and Firmicutes as shown in Figure 2. However, MEGAN also maintains a database with approximately 3,5000 viral genomes. Subject 7's plaque sample harbored 0.18% of viruses from the family *Retroviridae*, *Porcine type C oncovirus* strain PRJNA14126.

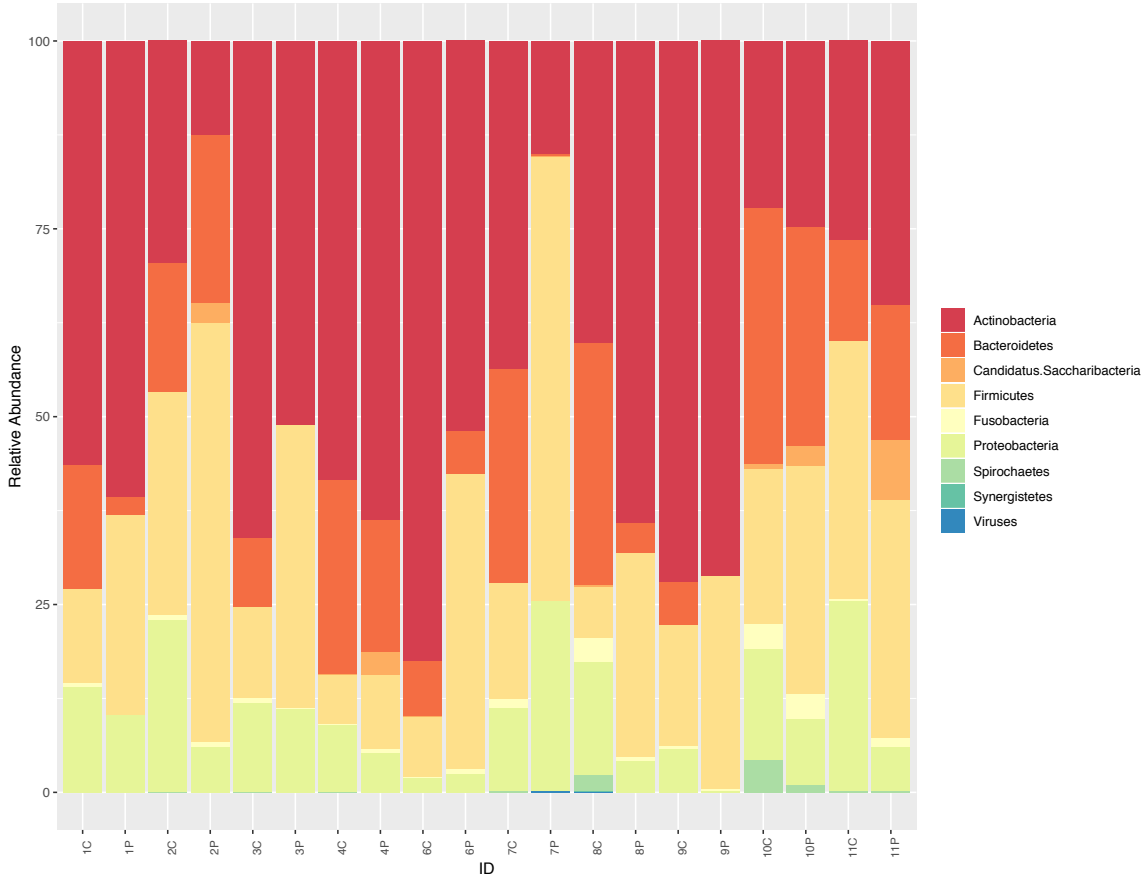


Figure 4.2: Dominant phyla across all samples as assigned by metaphlan2. Individuals are numbered 1-11 (there is no sample 5), marked C for calculus and P for plaque.

In species-level analysis in MEGAN, we observed 205 species. Dominant species in dental calculus include *Pseudopropionibacterium propionicum* (average 5.62%), *Lautropia mirabilis* (average 9.13%), *Corynebacterium durum* (average 6.88%), *Rothia sp. Olga* (average 8.03%), *Capnocytophaga gingivalis* (average 4.69%), and *Cornyebacterium matrouchotii* (average 8.21%). Species dominant in dental plaque include: *Cornyebacterium matrouchotii* (average 15.57%), *Gemella morbillorum* (average 2.53%), *Candidatus Saccharibacteria oral taxon TM7x* (average 3.57%), *Streptococcus sanguinus* (average 3.99%), and *Rothia dentocariosa* (average 13.58%).

We observed 132 species level identifications in metaphlan2. Dental calculus was dominated by *Streptococcus sanguinus* (average 11.34%), *Rothia dentocariosa* (average 11.97%), *Rothia aeria* (average 15.17%), *Captocytophaga gingivalis* (average average 3.9%), *Corynebacterium matrouchotti* (average 8.61%), and an unclassified species within *Capnocytophaga* (average 10.19%). Dominant species within dental plaque include *Corynebacterium matrouchotti* (average 12.03%), *Rothia dentocariosa* (average 20.99%), *Gemella morbillorum* (average 2.86%), *Veillonella parvula* (average 9.60%), *Streptococcus sanguinis* (average 7.94%), and an unclassified species within *Capnocytophaga* (average 3.7%).

Core microbiome

Core microbiota was computed at 50%, 75%, and 100% across dental calculus and plaque samples in MEGAN. Taxa present in 50% of all samples include: Bacteroidales, *Capnocytophaga*, *Neiserria*, Pasteurellaceae, *Veillonella*, *Actinomyces*, *Streptococcus*, *Corynebacterium matruchotii*, and *Rothia dentocariosa*. Taxa present in 75% of samples include: *Capnocytophaga*,

Neisseriaceae, Gammaproteobacteria, *Actinomyces*, *Corynebacterium*, *Rothia*, and *Streptococcus*. Taxa present in 100% of the samples are Gammaproteobacteria, *Actinomyces*, and *Streptococcus*.

Sourcetracking

We conducted two sourcetracking analyses. The first, with the subject's own plaque and their calculus samples in addition to other oral microhabitats derived from the Human Microbiome Project (HMP). The second line of analysis consisted of only with subject's calculus and all other oral microbiota were derived from HMP. Results for the first line of analysis is shown in Figure 4.3 and the second line of analysis is in Figure 4.4.

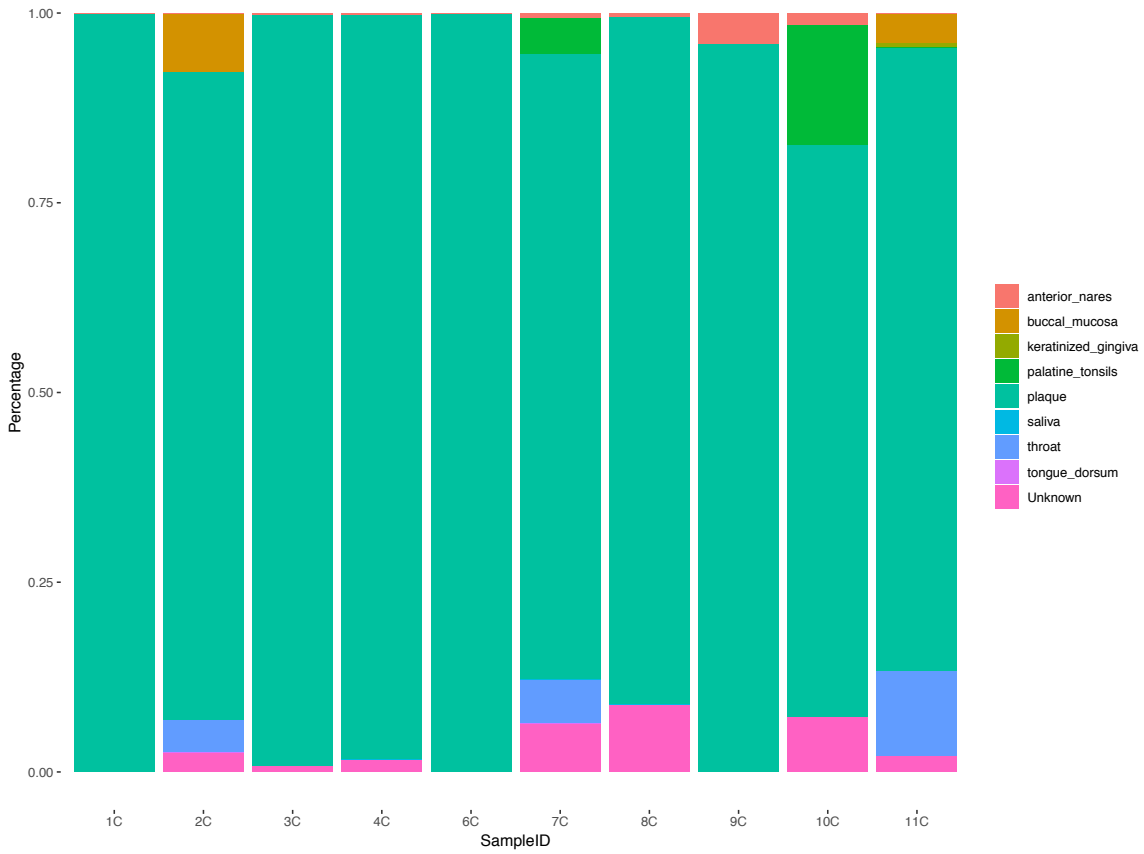


Figure 4.3: Sourcetracker results when comparing a subject's calculus (represented by the bar), to their plaque, and other oral microbial habitats downloaded from the HMP browser. Proportions of each microbial habitat are shown in colors on the legend.

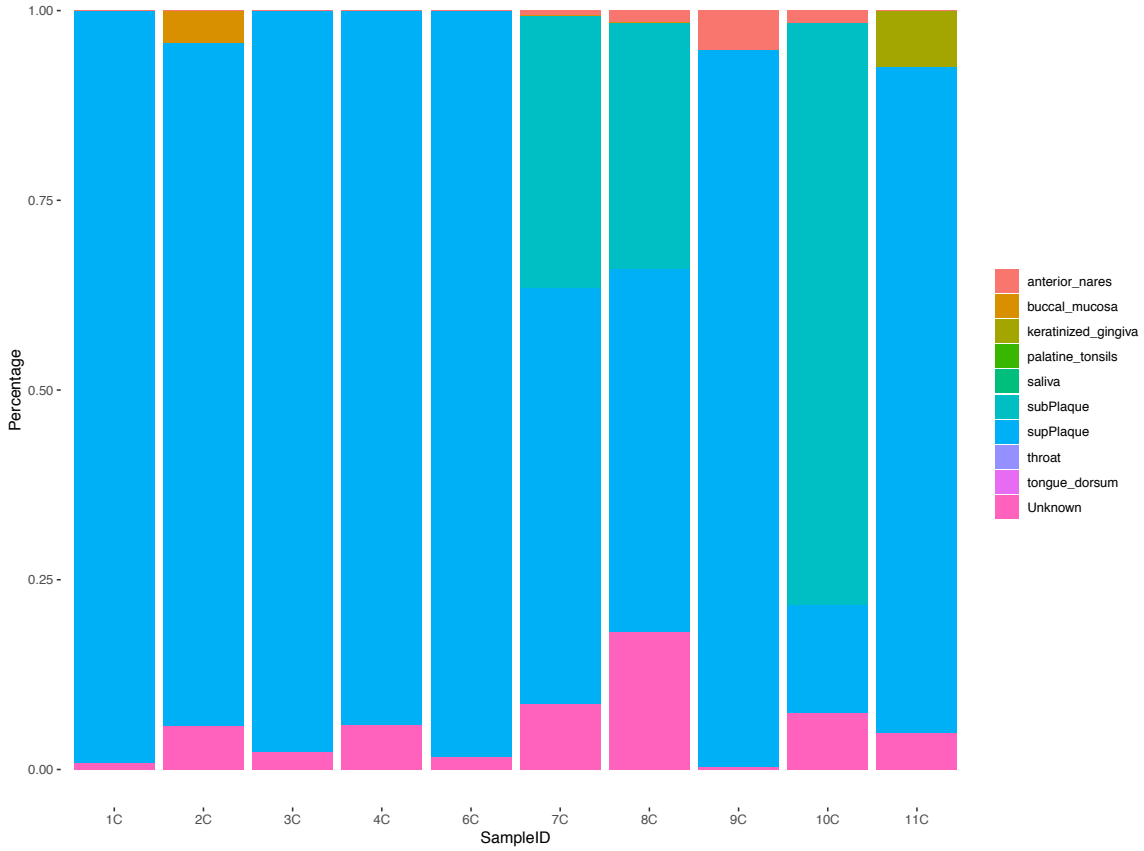


Figure 4.4: Sourcetracker results when comparing a subject's calculus (represented by the bar), to HMP supragingival and subgingival plaque as well as other oral microbial habitats downloaded from the HMP browser. Proportions of each microbial habitat are shown in colors on the legend.

In sourcetracking analysis using the subject's plaque and calculus samples, in addition to HMP oral microbiota, on average 90% of the individual's plaque contributed to their microbial communities in their calculus (range 75%-99%). Other microbial communities that contributed to calculus samples were throat microbiota, present on average 2% (0.00005%-11%), and anterior nares, present on average 0.006% (0.1%-4%). Buccal mucosa communities were present in 2 samples (3.9%-7.6%). Keratinized gingiva accounted for .6% of a single calculus sample. Palatine tonsil microflora contributed to 4-15% of two samples. Tongue dorsum and saliva microbial communities contributed the least to calculus samples, present on average 0.0001% and 0.000014% to calculus microbiota.

Upon comparing subjects' calculus to HMP-defined oral microbial communities, results showed calculus samples have high supragingival plaque contributions. We observed an average of 77.82% supragingival plaque contributions to every subject's calculus. In subject ten, this trend was flipped with 76.55% microbial contributions from subgingival plaque and 14.36% microbial contributions from supragingival plaque. Other oral microhabitats derived from HMP data contributed to calculus communities in our analysis. Anterior nares microbiota were present on average 0.89% (0.01%-5.5%). Buccal mucosa was present on average 0.42% (0.00003-4.2%. Keratinized gingiva microflora contributed to on average 3.7% (0.01%-7.3%) in two subjects. Saliva microbiota contributed to 0.000005% (0.000001%-0.000004). Tongue dorsum microbiota contributed 0.00005% to a single subjects calculus microbiota. Palatine tonsils and throat microbiota were absent in this analysis.

Indicator Species

In metaphlan2 analysis, out of the 132 species we analyzed, only 10 species had strong statistical support as indicator species for both calculus and

plaque. *Neiseria elongata* is the only indicator species for calculus (p=0.005). Other species with less strong support include *Cardiobacterium hominis* (p=0.020) and *Propionibacterium propionicum* (p=0.45). There were seven indicator species associated with dental plaque (of the same subject), these include *Gemella haemolysans*, *Streptococcus gordonii*, *Streptococcus mitis oralis pneumoniae*, and an unclassified species within *Veillonella* (all with p=0.005). The remaining three species were *Haemophilus haemolyticus*, *Granulicatella elegans*, and an unclassified member of *Gemella*.

We also did indicator analysis using species-level taxonomic assignments from MEGAN. Out of 205 species, 12 species were proposed as indicator species. However, all 12, with the exception of one species, lacked strong statistical support. *Veillonella parvula,* was the sole indicator species in this analysis, strongly associated with dental plaque (p=0.005).

Differentially Abundant Taxa

Using species-level assignments from metaphlan2, we found 22 species that were differentially abundant between calculus and plaque shown in Figure 4.5 (LDA>2). Thirteen species were more abundant in plaque. These species include *Veillionella parvula*, *Streptococcus gordonii*, *Streptococcus mutans*, *Streptococcus mitis oralis pneumoniae*, *Gemella hamolysans*, *Selenomonas noxia*, *Streptococcus salivarous*, *candidate division TM7 single isolate TM7b*, *Haemophilus influenzae*, *Haemophilus haemolyticus*, *Granulicatella elegans*, and unclassified members of *Veillionella* and *Gemella*. Nine species were more abundant in dental calculus. These include *Eubacterium yurii*, *Cardiobacterium hominis*, *Neisseria elongata*, *Bacteriodetes oral taxon 274*, *Propionibacterium*

propionicum, *Corynebacterium durum*, *Lautropia mirabilis*, *Capnoctyophaga gingivalis*, and *Rothia aeria*.

Species-level assignments from MEGAN were also input into LefSe, which detected 24 differentially abundant taxa between calculus and plaque shown in Figure 3 (LDA>2). Twelve species were differentially abundant in plaque. These included *Streptococcus gordonii*, *Veillonella parvula*, *Actinomyces sp. oral taxon 448*, *Streptococcus pnuemoniae*, *Selenomonas noxia*, *Streptoccoccus oralis*, *Granulicatella elegans*, *Kingella oralis*, *Gemella haemolysans*, *Veionella sp S13054-11*, *Streptococcus mitis*, and *Legionella pneumophila*. Twelve species were differentially abundant in dental calculus. These included *Cardiobacterium valvarum*, *Ottowia sp. oral taxon 894*, *Aggregatibacter aphrophilus*, *Eubacterium yurii, Capnocytophaga sputigena*, *Ottowia sp Marseille P4747*, *Cardiobacterium hominis*, *Neiserria elongata*, *Actinomyces sp Marsielle P3109*, *Corynebacterium durum*, *Pseudopropionibacterium propionicum*, and *Lautropia mirabilis*.

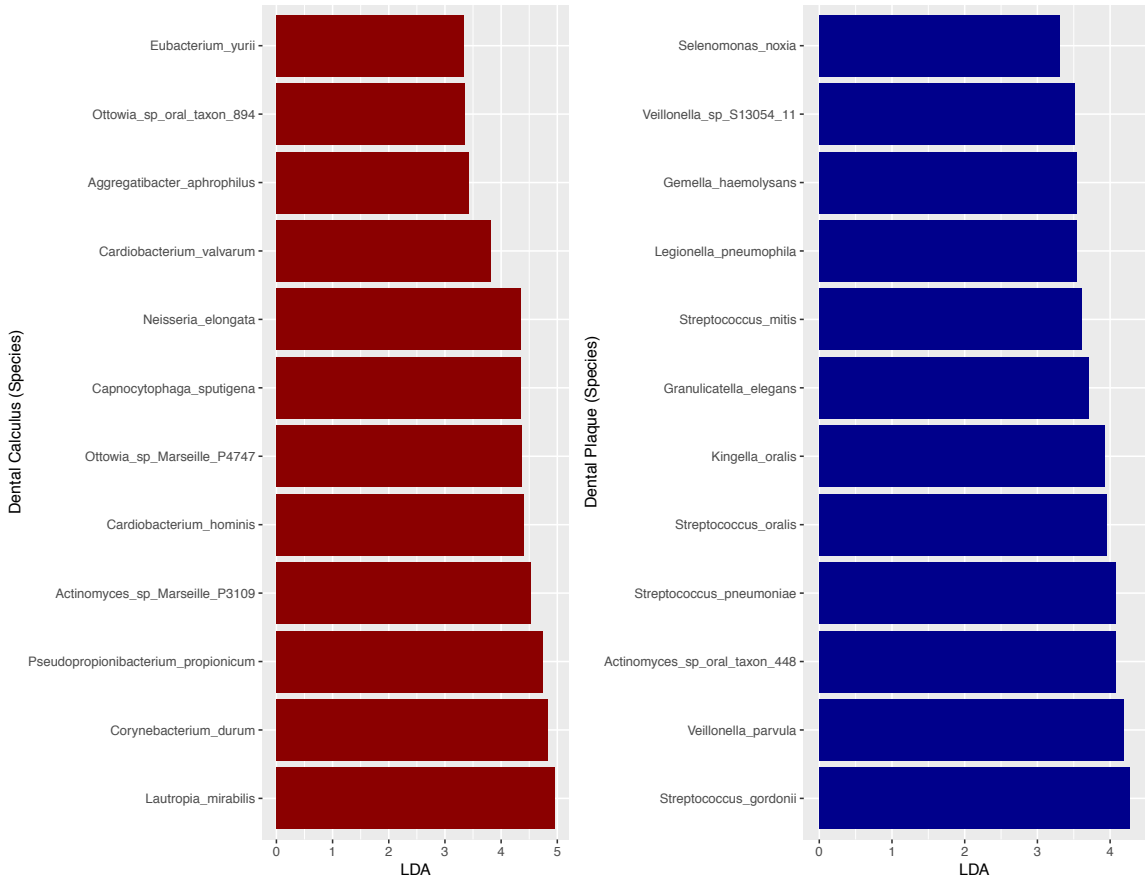


Figure 4.5: Results from Lefse analysis of differential species abundance comparison between calculus and plaque of the same individual.

Taxonomic Diversity

MEGAN species abundance files were used for diversity analysis due to their higher species-level assignments. Here we report observed alpha diversity (count of unique OTUs per sample) statistics, however all alpha diversity statistics by sample are available in Supplementary Tables 4.5 and 4.6. Observed estimated richness for subject's in this study dental calculus ranged from 106-225 taxa (average 152.1) and within dental plaque ranged from 97-195 taxa (average 143.2).

Alpha diversity analysis revealed that dental calculus is similar in terms of diversity to other HMP oral microbiota as shown in Figure 4.6. However, some of the highest OTU measurements were observed in dental calculus samples.

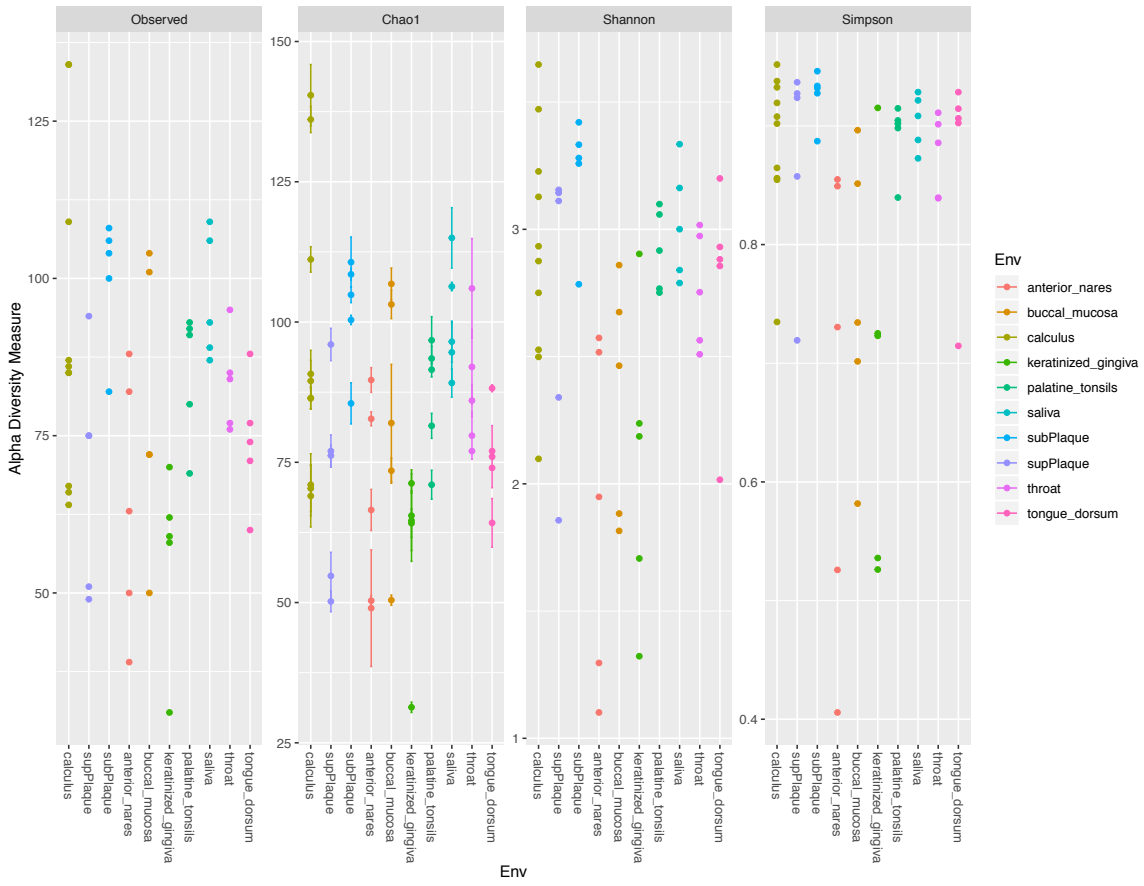


Figure 4.6: Alpha diversity of dental calculus (from subject's under study) compared to other oral microbial habitats (downloaded from the HMP browser) measured using Observed (count of unique OTUs per sample), Chao1 (estimation of diversity based on abundance data), Shannon's diversity index, and Simpson index constructed in phyloseq. Statistics are available in Supplementary Table 4.5

Alpha diversity of microbial species within dental calculus is expected to decline over time due to factors that led to the degradation and fragmentation of ancient DNA [179], however this assumption has not been tested. We compared these values to the alpha diversity of historical calculus samples (between 73-128 years older, Chapter 3) processed in a similar manner to modern calculus samples in Figure 4.7. While taxonomic diversity of modern plaque and modern calculus are similar, the taxonomic diversity of historic-era calculus can vary. Interestingly, the taxonomic diversity of observed unique taxa in historical calculus exceeded that of periodontal plaque metagenomes, which ranged from 118-183 taxa (average 153.9).

Observed richness estimated in historical calculus samples (analyzed in Chapter 3) ranged from 59-350 taxa (average 190.6), including both the highest and lowest values estimated. The lowest alpha diversity measurements were assigned to samples HTH399, HOW227864, MED98137_Q01, and KOD364459, three of which had Proteobacteria-dominant skin-like metagenomes, which suggests dysbiosis in the oral microbial communities may be reconstructed in historical calculus samples. Regardless, the historical samples do not show a clear trend of having lower alpha diversity than modern calculus or plaque samples.

While alpha diversity shows differences in the taxonomic diversity of historical calculus and other oral microenvironments, beta diversity analysis shows overlap in plaque and calculus metagenomes from this chapter. This is expected since plaque and calculus metagenomes were recovered from the same individual. Additionally, periodontal plaque metagenomes forms a distinct cluster while a majority of historical calculus metagenomes cluster away from all 3 communities, with a few exceptions. While a majority of individuals studied in

Chapter 3 had periodontitis (Sup table), their dental calculus metagenomes are clustering away from modern periodontal metagenomes.

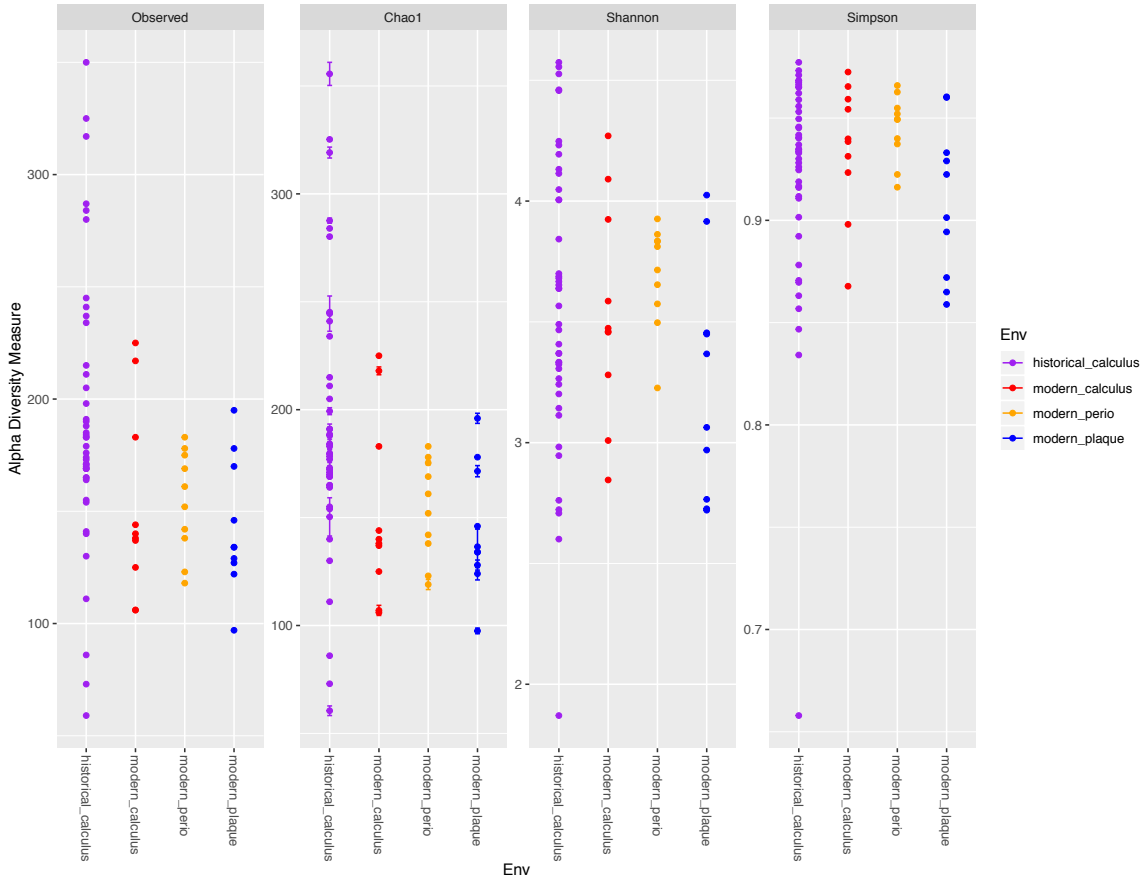


Figure 4.7: Alpha Diversity of historical calculus, modern calculus, modern periodontal plaque, and modern plaque metagenomes measured using Observed (count of unique taxa), Chao1 (estimation of diversity based on abundance data), Shannon's diversity index, and Simpson index constructed in phyloseq. Supplementary Table 4.6

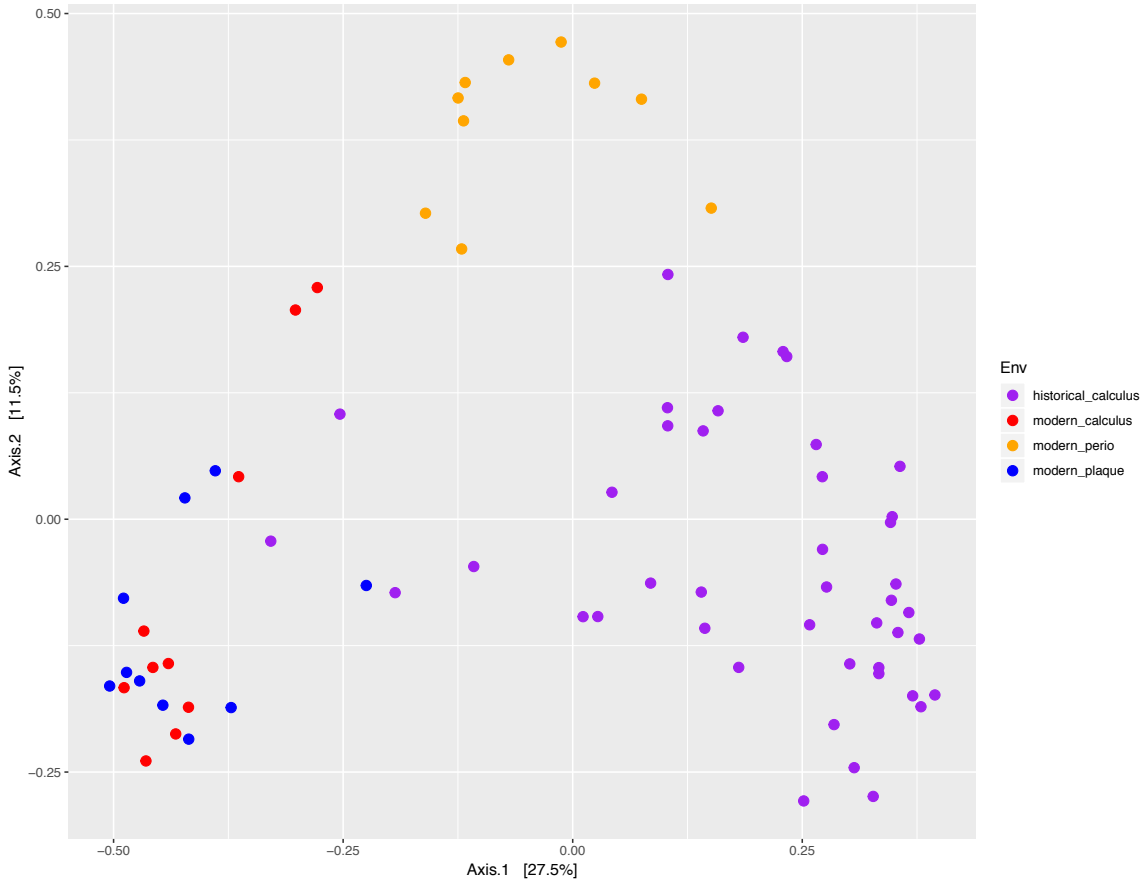


Figure 4.8: PCoA plot constructed using Bray-Curtis distances. Here we show the beta diversity of historical calculus (Chapter 3), modern calculus (Chapter 4), modern periodontal plaque, and modern plaque metagenomes (Chapter 4).

Discussion

In this study, we provided the first comprehensive analysis of dental calculus microbiota and compared the microbial composition of dental plaque and dental calculus. We found that while dental calculus resembles supragingival plaque, it is not an exact match. We observed microbial contributions from other oral microenvironments: buccal mucosa, palatine tonsils, the throat, and tongue dorsum to dental calculus.

Human-associated DNA differed between the two substrates, with dental plaque having higher amounts of human DNA as described in previous studies [127]. Human DNA in ancient calculus is low, ranging from 0.007%-0.47% [140]. Additionally historic-era calculus obtained from natural history museums, on average had 0.3% of human-associated DNA (Chapter 3). This suggests that, in life, dental calculus preserves less human associated DNA and that over time there will be even less to recover from the substrate. However, in-solution capture methods have been successful at targeting degraded and fragmented ancient nuclear and mitochondrial human DNA [180, 181].

On average, 90% of an individual's plaque microflora contributed to their calculus microflora. It is important to note, that sourcetracker is conducted at the genus level and, as suggested by beta diversity analysis, dominant species between plaque and calculus may greatly differ. Additionally, dental calculus is porous and able to absorb bacterial toxins secreted by periodontal pathogens and damage the tissue surrounding the periodontal pocket [16, 182, 183]. For these reasons, dental calculus is associated with gingival inflammation on the supragingival surface and affects the progression of periodontitis when present on the subgingival surface [183]. However, when our subject's calculus was compared to HMP supragingival and subgingival plaque communities, it

overlapped an average of 77.82% with supragingival plaque and less with subgingival plaque. Interestingly, historic-era dental calculus is mostly comprised of subgingival plaque, having on average 22.24% subgingival plaque, 0.9% supragingival plaque, and 61.66% of unknown microbial contributions. However, the overall low overlap with modern HMP communities indicates that the microbial ecology of plaque has likely changed significantly, even dramatically, in the last century.

Core microbiome analysis revealed similarities between dental plaque and calculus samples, with 100% of samples being comprised of members of *Actinomyces* and *Streptococcus*. This may be due to their location within dental calculus. *Actinomyces* and *Streptococcus* are constituents of early colonizers, which suggests later colonizers may be poorly preserved in ancient calculus. We observed few indicator species, with only eight having strong statistical support. *Neisseria elongata* was the only species associated with dental calculus with sufficient statistical support. Four species were associated with dental plaque: *Gemella haemolysans*, *Streptococcus gordonii*, *Streptococcus mitis oralis pneumoniae*, *Veillonella parvula*, and an unclassified species within *Veillonella*. This suggests, although bacteria belonging to *Streptococcus* and *Actinomyces*, are conserved between dental plaque and calculus, species within *Streptococcus* may be enriched in dental plaque and difficult to ascertain in ancient and historical dental calculus.

Differences between dental calculus and plaque microbiota were not found in taxa, but their abundance. We observed two unique differentially abundant species within dental calculus, *Neisseria elongata* and *Cardiobacterium hominis*, using different approaches. We also observed that 5 indicator species were also differentially abundant in their substrate. This

suggests reconstructing plaque indicator species within ancient calculus may be challenging.

Dental calculus has a distinct microbiome, but is most similar to supragingival plaque of the same individual or healthy supragingival plaque from published HMP datasets. Conducting alpha diversity analysis on ancient DNA is problematic [179]. However, when performed on modern calculus we found that dental calculus is as diverse and occasionally more diverse than other oral microhabitats. This diversity will likely decline over time and, when compounded with differentially abundant species, will bias findings. What remains unclear is whether the high alpha diversity of historical dental calculus samples is due to the lack of degradation in resident microflora or whether individuals who lived at the turn of the century had more diverse oral microbiomes. For this reason, it is important to distinguish dental calculus microbiome findings from general oral microbiome findings and not assume that calculus microbiota is a perfect reconstruction of an individual's plaque microbial communities.

Conclusion

Dental calculus is ubiquitous in ancient and modern living populations. Due to the enrichment of inorganic minerals, calculus preserves biomolecules of organisms that are thousands of years old. While the modern dental plaque microbiome is well-studied, the modern dental calculus microbiome is poorly understood. We conducted the first comparison of dental plaque and calculus communities. The dental calculus microbiome is a distinct microbial community with overlap in dental plaque microbiota. The differences lay in the abundance of microorganisms, including members of *Streptococcus*, *Veillonella*, *Gemella*, and *Neisseria*. Also while dental calculus as diverse as other microbial

communities, this diversity will decline over time, leading to a bias and misinterpretation of dental calculus microbiota as representative as a person's plaque microbiota.

References Cited

1. Putt, S.S., et al., *The functional brain networks that underlie Early Stone Age tool manufacture.* Nature Human Behaviour, 2017. 1(6): p. 0102.
2. Armelagos, G.J., *The paleolithic disease-scape, the hygiene hypothesis, and the second epidemiological transition*, in *The Hygiene Hypothesis and Darwinian Medicine*, G.A.W. Rook, Editor. 2009, Birkhäuser Basel: Basel. p. 29-43.
3. Omran, A.R., *A century of epidemiologic transition in the United States.* Preventive medicine, 1977. 6(1): p. 30-51.
4. Carrera-Bastos, P., et al., *The western diet and lifestyle and diseases of civilization.* Research Reports in Clinical Cardiology, 2011. 2(1): p. 15-35.
5. Omran, A.R., *The Epidemiologic Transition: A Theory of the Epidemiology of Population Change.* The Milbank Quarterly, 2005. 83(4): p. 731-757.
6. CDC/NCHS, N.V.S.S. *Leading Causes of Death: 1900-1998*. Available from: https://www.cdc.gov/nchs/data/dvs/lead1900_98.pdf.
7. Harper, K. and G. Armelagos, *The changing disease-scape in the third epidemiological transition.* Int J Environ Res Public Health, 2010. 7(2): p. 675-97.
8. Marthaler, T., *Changes in dental caries 1953–2003.* Caries research, 2004. 38(3): p. 173-181.
9. CDC/NCHS, N.V.S.S. *Dental, Ages 1-74 years, 4235*. NHANES I (1971-1974); Available from: https://wwwn.cdc.gov/nchs/data/nhanes1/4235.pdf.
10. NHANES. *Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention*. 2018; Available from: https://www.nidcr.nih.gov/research/data-statistics/dental-caries/adults.
11. Marsh, P.D., *Microbial Ecology of Dental Plaque and its Significance in Health and Disease.* Advances in Dental Research, 1994. 8(2): p. 263-271.
12. Ximénez-Fyvie, L.A., A.D. Haffajee, and S.S. Socransky, *Comparison of the microbiota of supra-and subgingival plaque in health and periodontitis.* Journal of clinical periodontology, 2000. 27(9): p. 648-657.
13. Hajishengallis, G., R.P. Darveau, and M.A. Curtis, *The keystone-pathogen hypothesis.* Nat Rev Microbiol, 2012. 10(10): p. 717-25.
14. Jiao, Y., M. Hasegawa, and N. Inohara, *The role of oral pathobionts in dysbiosis during periodontitis development.* Journal of dental research, 2014. 93(6): p. 539-546.
15. Han, Y.W. and X. Wang, *Mobile microbiome: oral bacteria in extra-oral infections and inflammation.* J Dent Res, 2013. 92(6): p. 485-91.
16. Marsh, P.D., et al., *Oral Microbiology*. 2009: Elsevier health sciences.
17. Haumschild, M.S. and R.J. Haumschild, *The Importance of Oral Health in Long-Term Care.* Journal of the American Medical Directors Association, 2009. 10(9): p. 667-671.
18. Lassalle, F., et al., *Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet.* Molecular Ecology, 2018. 27(1): p. 182-195.
19. Contreras, M., et al., *The bacterial microbiota in the oral mucosa of rural Amerindians.* Microbiology, 2010. 156(11): p. 3282-3287.
20. Clemente, J.C., et al., *The microbiome of uncontacted Amerindians.* Science advances, 2015. 1(3): p. e1500183.

21. Brealey, J.C., et al., *Dental calculus as a tool to study the evolution of the oral microbiome in mammals.* bioRxiv, 2019: p. 596791.
22. Adler, C.J., et al., *Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions.* Nat Genet, 2013. 45(4): p. 450-5, 455e1.
23. Santiago-Rodriguez, T.M., et al., *Commensal and Pathogenic Members of the Dental Calculus Microbiome of Badia Pozzeveri Individuals from the 11th to 19th Centuries.* Genes, 2019. 10(4): p. 299.
24. Santiago-Rodriguez, T.M., et al., *Insights of the dental calculi microbiome of pre-Columbian inhabitants from Puerto Rico.* PeerJ, 2017. 5: p. e3277.
25. Ziesemer, K.A., et al., *Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification.* Sci Rep, 2015. 5: p. 16498.
26. Shah, N., et al., *Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics.* Pac Symp Biocomput, 2011: p. 165-76.
27. Warinner, C., et al., *Pathogens and host immunity in the ancient human oral cavity.* Nat Genet, 2014. 46(4): p. 336-44.
28. Farrer, A.G., et al., *Biological and cultural drivers of oral microbiota in Medieval and Post-Medieval London, UK.* bioRxiv, 2018: p. 343889.
29. Marsh, P.D., *Are dental diseases examples of ecological catastrophes?* Microbiology, 2003. 149(Pt 2): p. 279-94.
30. Dahlen, G., *Microbiology and treatment of dental abscesses and periodontal-endodontic lesions.* Periodontology 2000, 2002. 28(1): p. 206-239.
31. Dörfer, C.E., et al., *The association of gingivitis and periodontitis with ischemic stroke.* Journal of Clinical Periodontology, 2004. 31(5): p. 396-401.
32. Kholy, K.E., R.J. Genco, and T.E. Van Dyke, *Oral infections and cardiovascular disease.* Trends in Endocrinology & Metabolism, 2015. 26(6): p. 315-321.
33. Scannapieco, F.A., R.B. Bush, and S. Paju, *Associations Between Periodontal Disease and Risk for Nosocomial Bacterial Pneumonia and Chronic Obstructive Pulmonary Disease. A Systematic Review.* Annals of Periodontology, 2003. 8(1): p. 54-69.
34. Detert, J., et al., *The association between rheumatoid arthritis and periodontal disease.* Arthritis Research & Therapy, 2010. 12(5): p. 218.
35. Abbayya, K., et al., *Association between Periodontitis and Alzheimer's Disease.* North American journal of medical sciences, 2015. 7(6): p. 241-246.
36. Castellarin, M., et al., *Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma.* Genome Research, 2012. 22(2): p. 299-306.
37. Harper, K. and G. Armelagos, *The changing disease-scape in the third epidemiological transition.* International Journal of Environmental Research and Public Health, 2010. 7(2): p. 675-97.
38. Pearlstein, K.E., *Health and the huddled masses: An analysis of immigrant and Euro-American skeletal health in 19 th century New York City*. 2015: American University.
39. Austin, R.M., et al., *Opinion: To curate the molecular past, museums need a carefully considered set of best practices.* Proceedings of the National Academy of Sciences, 2019. 116(5): p. 1471.
40. Hrdlička, A., *Biographical Memoir of George Sumner Huntington (1861-1927)*. 1937: National Academy of Sciences.

41. Betz, B.J., *Pastoralism, Agriculture, and Stress: A Comparative Analysis of Two 19th Century Qing Dynasty Populations*. 2013, The Ohio State University.
42. Voss, B.L. and R. Allen, *Overseas Chinese archaeology: Historical foundations, current reflections, and new directions.* Historical Archaeology, 2008. 42(3): p. 5-28.
43. Friday, C., *Organizing Asian-American Labor: The Pacific Coast Canned-Salmon Industry, 1870-1942*. 2010: Temple University Press.
44. Hrdlička, A., *The Anthropology of Kodiak Island*. 1944: Wistar Institute of Anatomy and Biology.
45. Henry, J., et al., *Reckoning With the Dead--The Larsen Bay Repatriation and the Smithsonian Institution.* 1994.
46. Muller, J.L., K.E. Pearlstein, and C. de la Cova, *Dissection and documented skeletal collections: Embodiments of legalized inequality*, in *The bioarchaeology of dissection and autopsy in the United States*. 2017, Springer. p. 185-201.
47. Cobb, W.M., *Human archives (Doctoral dissertation, Western Reserve University)*. 1932.
48. Todd, T.W., *The nature of mummification and maceration illustrated by the male white skull.* Journal of anatomy, 1925. 59(Pt 2): p. 180.
49. Hillson, S., *Recording dental caries in archaeological human remains.* International Journal of Osteoarchaeology, 2001. 11(4): p. 249-289.
50. Smith, B.H., *Patterns of molar wear in hunter–gatherers and agriculturalists.* American Journal of Physical Anthropology, 1984. 63(1): p. 39-56.
51. Greene, T.R., C.L. Kuba, and J.D. Irish, *Quantifying calculus: A suggested new approach for recording an important indicator of diet and dental health.* HOMO - Journal of Comparative Human Biology, 2005. 56(2): p. 119-132.
52. Dias, G. and N. Tayles, *'Abscess cavity'—a misnomer.* International Journal of Osteoarchaeology, 1998. 7(5): p. 548-554.
53. RStudio, T. *RStudio: integrated development for R*. 2015 [cited 42; Available from: http://www.rstudio.com.
54. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2016: Springer.
55. Lans, A., *"Whatever Was Once Associated with him, Continues to Bear his Stamp": Articulating and Dissecting George S. Huntington and His Anatomical Collection*, in *Bioarchaeological Analyses and Bodies*. 2018, Springer. p. 11-26.
56. Komar, D.A. and C. Grivas, *Manufactured populations: What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection.* American Journal of Physical Anthropology, 2008. 137(2): p. 224-233.
57. Newbrun, E., *Cariology 3 Ed*. 1975.
58. Phillips, S.M., *Inmate life in the Oneida County Asylum, 1860-1895: A biocultural study of the skeletal and documentary records*. 2001, State University of New York at Albany.
59. Geber, J. and E. Murphy, *Dental markers of poverty: Biocultural deliberations on oral health of the poor in mid-nineteenth-century Ireland.* American journal of physical anthropology, 2018. 167(4): p. 840-855.

60. Kashket, S., J. Zhang, and J. Van Houte, *Accumulation of fermentable sugars and metabolic acids in food particles that become entrapped on the dentition.* Journal of dental research, 1996. 75(11): p. 1885-1891.
61. Wols, H.D. and J.E. Baker, *Dental health of elderly confederate veterans: evidence from the Texas State Cemetery.* Am J Phys Anthropol, 2004. 124(1): p. 59-72.
62. Buzon, M.R., et al., *Health and disease in nineteenth-century San Francisco: skeletal evidence from a forgotten cemetery.* Historical archaeology, 2005. 39(2): p. 1-15.
63. Little, B.J., K.M. Lanphear, and D.W. Owsley, *Mortuary Display and Status in a Nineteenth-Century Anglo-American Cemetery in Manassas, Virginia.* American Antiquity, 1992. 57(3): p. 397-418.
64. Botha, D. and M. Steyn, *Dental health of the late 19th and early 20th century Khoesan.* HOMO-Journal of Comparative Human Biology, 2015. 66(3): p. 187-202.
65. Mandel, I.D., *Biochemical aspects of calculus formation: II. Comparative Studies of Saliva in Heavy and Light Calculus Formers.* Journal of Periodontal Research, 1974. 9(4): p. 211-221.
66. (ADA), A.D.A. *ADA History Timeline*. [cited 2018 December]; Available from: https://www.ada.org/en/about-the-ada/ada-history-and-presidents-of-the-ada/ada-timeline.
67. Wasterlain, S.N., S. Hillson, and E. Cunha, *Dental caries in a Portuguese identified skeletal sample from the late 19th and early 20th centuries.* American Journal of Physical Anthropology, 2009. 140(1): p. 64-79.
68. Phillips, S.M., *Inmate life in the Oneida County Asylum, 1860-1895: A biocultural study of the skeletal and documentary records (Doctoral dissertation, State University of New York at Albany)*. 2001.
69. (CDC), C.f.D.C.a.P. *Dental Caries (Tooth Decay) in Adults (Age 20 to 64)*. 2004 [cited 2018 December]; Available from: https://www.nidcr.nih.gov/research/data-statistics/dental-caries/adults.
70. Rathore, M., A. Singh, and V.A. Pant, *The Dental Amalgam Toxicity Fear: A Myth or Actuality.* Toxicology International, 2012. 19(2): p. 81-88.
71. Bödecker, C.F., *The modified dental caries index.* Journal of the American Dental Association, 1939. 26: p. 1453-1460.
72. Bödecker, C.F. and H. Bödecker, *A practical index of the varying susceptibility to dental caries in man.* J Dent Cosmos, 1931. 77: p. 707-16.
73. Broadbent, J.M. and W.M. Thomson, *For debate: problems with the DMF index pertinent to dental caries data analysis.* Community dentistry and oral epidemiology, 2005. 33(6): p. 400-409.
74. Wood, J.W., et al., *The osteological paradox: problems of inferring prehistoric health from skeletal samples [and comments and reply].* 1992. 33(4): p. 343-370.
75. Farmer, P., *Social inequalities and emerging infectious diseases.* Emerging Infectious Diseases, 1996. 2(4): p. 259.
76. Armelagos, G.J., P.J. Brown, and B. Turner, *Evolutionary, historical and political economic perspectives on health and disease.* Social Science & Medicine, 2005. 61(4): p. 755-765.
77. Gravlee, C.C., *How race becomes biology: embodiment of social inequality.* American journal of physical anthropology, 2009. 139(1): p. 47-57.
78. Acevedo-Garcia, D., *Residential segregation and the epidemiology of infectious diseases.* Social science & medicine, 2000. 51(8): p. 1143-1161.

79. Reda, S.F., et al., *Inequality in utilization of dental services: a systematic review and meta-analysis.* American journal of public health, 2018. 108(2): p. e1-e7.
80. Yaussy, S.L. and S.N. DeWitte, *Calculus and survivorship in medieval London: The association between dental disease and a demographic measure of general health.* Am J Phys Anthropol, 2019.
81. Innes, N., et al., *A Century of Change towards Prevention and Minimal Intervention in Cariology.* Journal of dental research, 2019. 98(6): p. 611-617.
82. Skelly, E., et al., *Consequences of colonialism: A microbial perspective to contemporary Indigenous health.* Am J Phys Anthropol, 2018. 167(2): p. 423-437.
83. Cho, I. and M.J. Blaser, *The human microbiome: at the interface of health and disease.* Nat Rev Genet, 2012. 13(4): p. 260-70.
84. Rook, G.A. and L.R. Brunet, *Microbes, immunoregulation, and the gut.* Gut, 2005. 54(3): p. 317-20.
85. Rook, G.A., C.L. Raison, and C.A. Lowry, *Microbial 'old friends', immunoregulation and socioeconomic status.* Clin Exp Immunol, 2014. 177(1): p. 1-12.
86. Dye, B.A., et al., *Dental caries and tooth loss in adults in the United States, 2011-2012*. 2015: US Department of Health and Human Services, Centers for Disease Control and ….
87. Eke, P.I., et al., *Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012.* Journal of periodontology, 2015. 86(5): p. 611-622.
88. Wade, W.G., *The oral microbiome in health and disease.* Pharmacological research, 2013. 69(1): p. 137-143.
89. Kerr, A.R., *The oral microbiome and cancer.* American Dental Hygienists' Association, 2015. 89(suppl 1): p. 20-23.
90. Kudo, Y., et al., *Oral environment and cancer.* Genes and Environment, 2016. 38(1): p. 13.
91. Whitmore, S.E. and R.J. Lamont, *Oral bacteria and cancer.* PLoS pathogens, 2014. 10(3): p. e1003933.
92. Knox, K.W. and N. Hunter, *The role of oral bacteria in the pathogenesis of infective endocarditis.* Aust Dent J, 1991. 36(4): p. 286-92.
93. Castellarin, M., et al., *Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma.* Genome Res, 2012. 22(2): p. 299-306.
94. Dominy, S.S., et al., *Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors.* Science advances, 2019. 5(1): p. eaau3333.
95. Sonnenburg, E.D. and J.L. Sonnenburg, *The ancestral and industrialized gut microbiota and implications for human health.* Nature Reviews Microbiology, 2019. 17(6): p. 383.
96. Weyrich, L., et al., *Laboratory contamination over time during low-biomass sample analysis.* bioRxiv, 2018: p. 460212.
97. Lindgreen, S., *AdapterRemoval: easy cleaning of next-generation sequencing reads.* BMC research notes, 2012. 5(1): p. 337.
98. Truong, D.T., et al., *MetaPhlAn2 for enhanced metagenomic taxonomic profiling.* 2015. 12(10): p. 902.
99. Afshinnekoo, E., et al., *Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics.* Cell Syst, 2015. 1(1): p. 72-87.

100. Li, H., et al., *The sequence alignment/map format and SAMtools.* 2009. 25(16): p. 2078-2079.
101. Jónsson, H., et al., *mapDamage2. 0: fast approximate Bayesian estimates of ancient DNA damage parameters.* Bioinformatics, 2013. 29(13): p. 1682-1684.
102. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nature methods, 2010. 7(5): p. 335.
103. Knights, D., et al., *Bayesian community-wide culture-independent microbial source tracking.* Nature methods, 2011. 8(9): p. 761.
104. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nature methods, 2015. 12(1): p. 59.
105. Huson, D.H., et al., *MEGAN analysis of metagenomic data.* Genome Res, 2007. 17(3): p. 377-86.
106. Davis, N.M., et al., *Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data.* Microbiome, 2018. 6(1): p. 226.
107. Zhang, J., et al., *PEAR: a fast and accurate Illumina Paired-End reAd mergeR.* 2013. 30(5): p. 614-620.
108. Willmann, C., et al., *Oral health status in historic population: Macroscopic and metagenomic evidence.* PLoS One, 2018. 13(5): p. e0196482.
109. Li, D., et al., *MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.* 2015. 31(10): p. 1674-1676.
110. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* 2013. 29(8): p. 1072-1075.
111. Wattam, A.R., et al., *PATRIC, the bacterial bioinformatics database and analysis resource.* 2013. 42(D1): p. D581-D591.
112. Darling, A.C., et al., *Mauve: multiple alignment of conserved genomic sequence with rearrangements.* 2004. 14(7): p. 1394-1403.
113. Aziz, R.K., et al., *The RAST Server: rapid annotations using subsystems technology.* 2008. 9(1): p. 75.
114. Alikhan, N.-F., et al., *BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons.* 2011. 12(1): p. 1.
115. Albanese, D. and C.J.N.c. Donati, *Strain profiling and epidemiology of bacterial species from metagenomic sequencing.* 2017. 8(1): p. 2260.
116. Zhou, Y., et al., *PHAST: a fast phage search tool.* Nucleic acids research, 2011. 39(suppl_2): p. W347-W352.
117. Gu, Z., *Complexheatmap: Making complex heatmaps. r package version 1.7. 1.* Reference Source.
118. De Caceres, M., F. Jansen, and M.M. De Caceres, *Package 'indicspecies'.* Relationship between species and groups of sites. R package version, 2016. 1(6).
119. Griffith, D.M., J.A. Veech, and C.J. Marsh, *Cooccur: probabilistic species co-occurrence analysis in R.* Journal of Statistical Software, 2016. 69(2): p. 1-17.
120. Oh, J., et al., *Temporal stability of the human skin microbiome.* 2016. 165(4): p. 854-866.
121. Huttenhower, C., et al., *Structure, function and diversity of the healthy human microbiome.* 2012. 486(7402): p. 207.
122. Schnorr, S., et al., *Gut microbiome of the Hadza hunter-gatherers. Nat Commun 5: 3654*. 2014.

123. Obregon-Tito, A.J., et al., *Subsistence strategies in traditional societies distinguish gut microbiomes.* 2015. 6: p. 6505.
124. Johnston, E.R., et al., *Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem.* 2016. 7: p. 579.
125. Velsko, I.M., et al., *Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research.* MSystems, 2018. 3(4): p. e00080-18.
126. Cobb, W.M., *Human archives*. 1932, Western Reserve University.
127. Liu, B., et al., *Deep sequencing of the oral microbiome reveals signatures of periodontal disease.* PLoS One, 2012. 7(6): p. e37919.
128. Ferrari, A., et al., *Isolation and characterization ofMethanobrevibacter oralis sp. nov.* Current Microbiology, 1994. 29(1): p. 7-12.
129. Khelaifia, S., et al., *Draft Genome Sequencing of Methanobrevibacter oralis Strain JMR01, Isolated from the Human Intestinal Microbiota.* Genome Announc, 2014. 2(1).
130. Lepp, P.W., et al., *Methanogenic Archaea and human periodontal disease.* Proc Natl Acad Sci U S A, 2004. 101(16): p. 6176-81.
131. Robichaux, M., M. Howell, and R. Boopathy, *Methanogenic activity in human periodontal pocket.* Current microbiology, 2003. 46(1): p. 0053-0058.
132. Lurie-Weinberger, M.N. and U. Gophna, *Archaea in and on the human body: health implications and future directions.* PLoS pathogens, 2015. 11(6): p. e1004833.
133. Socransky, S., et al., *Microbial complexes in subgingival plaque.* Journal of clinical periodontology, 1998. 25(2): p. 134-144.
134. *Anaerolineaceae*, in *Bergey's Manual of Systematics of Archaea and Bacteria*.
135. Shao, T., et al., *Combined Signature of the Fecal Microbiome and Metabolome in Patients with Gout.* Frontiers in Microbiology, 2017. 8(268).
136. Chang, C., et al., *The prevalence rate of periodontal pathogens and its association with oral squamous cell carcinoma.* Applied Microbiology and Biotechnology, 2019. 103(3): p. 1393-1404.
137. Cross, K.L., et al., *Insights into the Evolution of Host Association through the Isolation and Characterization of a Novel Human Periodontal Pathobiont, Desulfobulbus oralis.* mBio, 2018. 9(2): p. e02061-17.
138. Dridi, B., et al., *High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol.* PloS one, 2009. 4(9): p. e7063.
139. Raymann, K., et al., *Unexplored Archaeal Diversity in the Great Ape Gut Microbiome.* mSphere, 2017. 2(1): p. e00026-17.
140. Mann, A.E., et al., *Differential preservation of endogenous human and microbial DNA in dental calculus and dentin.* Sci Rep, 2018. 8(1): p. 9822.
141. Popejoy, A.B. and S.M. Fullerton, *Genomics is failing on diversity.* Nature News, 2016. 538(7624): p. 161.
142. Ozga, A.T., et al., *Oral microbiome diversity among Cheyenne and Arapaho individuals from Oklahoma.* American journal of physical anthropology, 2016. 161(2): p. 321-327.
143. Horz, H.P., *Archaeal Lineages within the Human Microbiome: Absent, Rare or Elusive?* Life (Basel), 2015. 5(2): p. 1333-45.

144. Li, C., et al., *Prevalence and molecular diversity of Archaea in subgingival pockets of periodontitis patients.* Oral microbiology and immunology, 2009. 24(4): p. 343-346.
145. Yamabe, K., et al., *Distribution of Archaea in Japanese patients with periodontitis and humoral immune response to the components.* FEMS microbiology letters, 2008. 287(1): p. 69-75.
146. Vianna, M., et al., *Quantitative analysis of three hydrogenotrophic microbial groups, methanogenic archaea, sulfate-reducing bacteria, and acetogenic bacteria, within plaque biofilms associated with human periodontal disease.* Journal of bacteriology, 2008. 190(10): p. 3779-3785.
147. Huynh, H.T.T., et al., *Restricted diversity of dental calculus methanogens over five centuries, France.* Scientific Reports, 2016. 6: p. 25775.
148. Zhang, Q., et al., *CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes.* Genome Biology, 2013. 14(4): p. R40.
149. Baker, J.L., et al., *Ecology of the oral microbiome: beyond bacteria.* Trends in microbiology, 2017. 25(5): p. 362-374.
150. Abeles, S.R., et al., *Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes.* PLoS One, 2015. 10(8): p. e0134941.
151. Robles-Sikisaka, R., et al., *Association between living environment and human oral viral ecology.* ISME J, 2013. 7(9): p. 1710-24.
152. Pride, D.T., et al., *Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome.* The Isme Journal, 2011. 6: p. 915.
153. Ly, M., et al., *Altered oral viral ecology in association with periodontal disease.* MBio, 2014. 5(3): p. e01133-14.
154. Kiple, K.F. and K.C. Ornelas, *The Cambridge World History of Food: Volume One*. 2015: Cambridge University Press.
155. De, S., *Dressed vegetables à la mode*. 1892, London: Longmans. 85 p.
156. Roundell, C., *Mrs. Roundell's practical cookery book: with many family recipes hitherto unpublished*. Practical cookery book. 1898, London: Bickers. viii, 580 p.
157. Senn, C.H., *How to cook vegetables*. 1911, London: The Food and cookery publishing agency. 220 p.
158. Bellows, A.J., *The philosophy of eating*. 1867, New York: Hurd and Houghton; Boston, E. P. Dutton and company. 2 p. l., 3-342 p.
159. Ward, A., *The Grocer's Encyclopedia*. 1911: A. Ward.
160. Sklivaniotis, M. and A. Angelakis, *Water for Human Consumption through the History*. 2019.
161. Cutler, D. and G. Miller, *The role of public health improvements in health advances: The twentieth-century United States.* Demography, 2005. 42(1): p. 1-22.
162. Duffy, J., *The sanitarians: a history of American public health*. 1992: University of Illinois Press.
163. Dijkshoorn, L., A. Nemec, and H. Seifert, *An increasing threat in hospitals: multidrug-resistant Acinetobacter baumannii.* Nat Rev Microbiol, 2007. 5(12): p. 939-51.
164. Peleg, A.Y., H. Seifert, and D.L. Paterson, *Acinetobacter baumannii: emergence of a successful pathogen.* Clin Microbiol Rev, 2008. 21(3): p. 538-82.

165. Maragakis, Lisa L. and Trish M. Perl, *Antimicrobial Resistance:Acinetobacter baumannii:Epidemiology, Antimicrobial Resistance, and Treatment Options.* Clinical Infectious Diseases, 2008. 46(8): p. 1254-1263.
166. Fournier, P.-E., et al., *Comparative genomics of multidrug resistance in Acinetobacter baumannii.* PLoS genetics, 2006. 2(1): p. e7.
167. Benjamin, A.E., H. Ahmed, and G.B.A. Sebastian, *The Rise of Carbapenem-Resistant Acinetobacter baumannii.* Current Pharmaceutical Design, 2013. 19(2): p. 223-238.
168. Brooke, J.S., *Stenotrophomonas maltophilia: an emerging global opportunistic pathogen.* Clin Microbiol Rev, 2012. 25(1): p. 2-41.
169. Hershkovitz, I., et al., *Oral bacteria in Miocene Sivapithecus.* Journal of human evolution, 1997. 33(4): p. 507-512.
170. Henry, A.G., et al., *The diet of Australopithecus sediba.* Nature, 2012. 487(7405): p. 90.
171. Weyrich, L.S., et al., *Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus.* Nature, 2017.
172. Marsh, P., *Dental plaque as a microbial biofilm.* Caries research, 2004. 38(3): p. 204-211.
173. Valm, A.M., *The Structure of Dental Plaque Microbial Communities in the Transition from Health to Dental Caries and Periodontal Disease.* Journal of molecular biology, 2019.
174. Lieverse, A.R., *Diet and the aetiology of dental calculus.* International Journal of osteoarchaeology, 1999. 9(4): p. 219-232.
175. Warinner, C., C. Speller, and M.J. Collins, *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, 2015. 370(1660): p. 20130376.
176. Segata, N., et al., *Metagenomic biomarker discovery and explanation.* 2011. 12(6): p. R60.
177. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data.* PloS one, 2013. 8(4): p. e61217.
178. Allaire, J., *RStudio: integrated development environment for R.* Boston, MA, 2012.
179. Dabney, J., M. Meyer, and S. Pääbo, *Ancient DNA damage.* Cold Spring Harbor perspectives in biology, 2013. 5(7): p. a012567.
180. Ozga, A.T., et al., *Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus.* Am J Phys Anthropol, 2016. 160(2): p. 220-8.
181. Ziesemer, K.A., et al., *The efficacy of whole human genome capture on ancient dental calculus and dentin.* American Journal of Physical Anthropology, 2019. 168(3): p. 496-509.
182. Kamath, D.G. and S. Umesh Nayak, *Detection, removal and prevention of calculus: Literature Review.* Saudi Dent J, 2014. 26(1): p. 7-13.
183. Mandel, I.D., *Calculus update: prevalence, pathogenicity and prevention.* The Journal of the American Dental Association, 1995. 126(5): p. 573-580.

Score	Dental Site	NMNH-Huntington		NMNH-Howard		NMNH-Kodiak		CMNH-HamannTodd	
		%	No./Total	%	No./Total	%	No./Total	%	No./Total
0	Interproximal	18%	9/50	43%	85/200	29%	57/196	27%	102/217
	Lingual	43%	22/51	54%	107/200	47%	92/196	52%	121/233
	Labial/Buccal	29%	15/51	52%	105/200	51%	100/196	69%	157/226
1	Interproximal	42%	21/50	41%	81/200	35%	69/196	35%	75/217
	Lingual	37%	19/51	32%	63/200	21%	42/196	26%	60/233
	Labial/Buccal	22%	11/51	29%	58/200	26%	51/196	12%	27/226
2	Interproximal	16%	8/50	14%	27/200	16%	31/196	10%	21/217
	Lingual	6%	3/51	8%	16/200	13%	26/196	12%	28/233
	Labial/Buccal	16%	8/51	5%	9/200	8%	16/196	8%	18/226
3	Interproximal	24%	12/50	4%	7/200	20%	39/196	9%	19/217
	Lingual	14%	7/51	7%	14/200	18%	36/196	10%	24/233
	Labial/Buccal	33%	17/51	14%	28/200	15%	29/196	11%	24/226
Total 0	All Dental Sites	28%	46/152	50%	297/600	42%	249/588	56%	380/676
Total 1	All Dental Sites	34%	51/152	34%	202/600	36%	211/588	24%	162/676
Total 2	All Dental Sites	13%	19/152	9%	52/600	12%	73/588	10%	67/676
Total 3	All Dental Sites	24%	36/152	8%	49/600	18%	104/588	10%	67/676
Total 1-3	All Dental Sites	70%	106/152	50%	303/600	66%	388/588	44%	296/676

Supplementary Table 2.3: By collection, the dental calculus scores given according to Greene et al. (2005) are shown by percentages evaluated in tooth as well as the overall number and the total sites observed.

Score	NMNH-Huntington		NMNH-Howard		NMNH-Kodiak		CMNH-HamannTodd	
	%	No./Total	%	No./Total	%	No./Total	%	No./Total
0	0%	0/49	0%	0/184	1.7%	3/185	0%	0/211
1	57%	28/49	52%	96/184	30%	56/185	37%	78/211
2	18%	9/49	23%	42/184	28%	52/185	32%	68/211
3	16%	8/49	15%	27/184	23%	43/185	18%	38/211
4	6%	3/49	6%	12/184	8%	15/185	8%	17/211
5	2%	1/49	3%	5/184	6%	12/185	4%	9/211
6	0%	0/49	0.5%	1/184	2%	4/185	.4%	1/211
Totals 1-2	76%	37/49	75%	138/184	58%	108/185	69%	146/211
Totals 3-4	22%	11/49	20%	39/184	31%	58/185	26%	55/211
Totals 5-6	2%	1/49	3%	6/184	9%	16/185	5%	10/211
Totals 1-6	100%	49/49	99%	183/184	98%	182/185	100%	211/211

Supplementary Table 2.4: By collection, the attrition scores given according to Smith (1984) are shown by percentages evaluated in tooth as well as the overall number and the total sites observed.

Time Period	N	Dataset	Location	Caries	Calculus	Attrition	AMTL	Enamel defects	Abscesses	Fillings	Reference
This study 1890-1920CE	1,123	Historical Skeletal Collections	Across the USA	22.1%	54.2%	99%	15.8%	26%	41%	Y	Williams et al. 2019 (Chapter 2)
1895-1928CE	9.562	Skeletal Collection	Portugal	27.9%[a]	N/A	N/A	37.2%	N/A	N/A	Y	Wasterlain et al. 2009
1821-1874CE	8,832	Church Cemetery	Belleville, Ontario	21.6%	N/A	N/A	23.6%	N/A	N/A	N/A	Saunders et al. 1997
1907-1932CE	1,366	Burial Site of Confederate Veterans	Austin, TX	24.4%	N/A	N/A	57.2%	8.2%	1.4%	Y	Wols and Baker 2004
Late 1800s-early 1900s CE	1,722	Skeletal Collection of Khoesan	Cape Colony, Southern Africa	6.5%	44%[d]	3.8%[c]	11.9%[b]	N/A	29.3%	N/A	Botha and Steyn 2015
1847-1851CE	6,860	Former Inmates of workhouse	Ireland	79.9%[e]	97.3%[e]	N/A	64.0%[e]	21.8%[e]	26.8%[e]	N/A	Geber and Murphy 2018
1830s-1907CE	197	Family Plantation	Manassas, VA	35%	N/A	N/A	N/A	7%	N/A	Y	Little et al. 1992
1860-1895CE	1,708	Oneida Country Asylum Inmates	Oneida, NY	36.1%	N/A	N/A	53.8%	N/A	1%	Y	Phillips 2001
1860-1895CE	606	Almshouse	Albany, NY	16.9%	N/A	N/A	38.6%	N/A	1.2%	N	Phillips 2001
Mid-19th century	80[d]	Cemetery	San Francisco, CA	45%[d]	N/A	N/A	80%[d]	50%[d]	34%[d]	N/A	Buzon et al. 2005

Supplementary Table 2.5: Cross-study dental paleopathology comparisons. This table lists the time period under study, the total number of teeth examined, a description of the dataset analyzed, the location where the data was derived, and prevalence for caries, calculus, attrition, antemortem tooth loss, enamel defects, and abscesses. Fillings were marked "Y" if fillings were observed, described, and found in the study and marked "N" if fillings were observed, described, and not found in the study. "N/A" denotes the pathology was not described in the specified study.

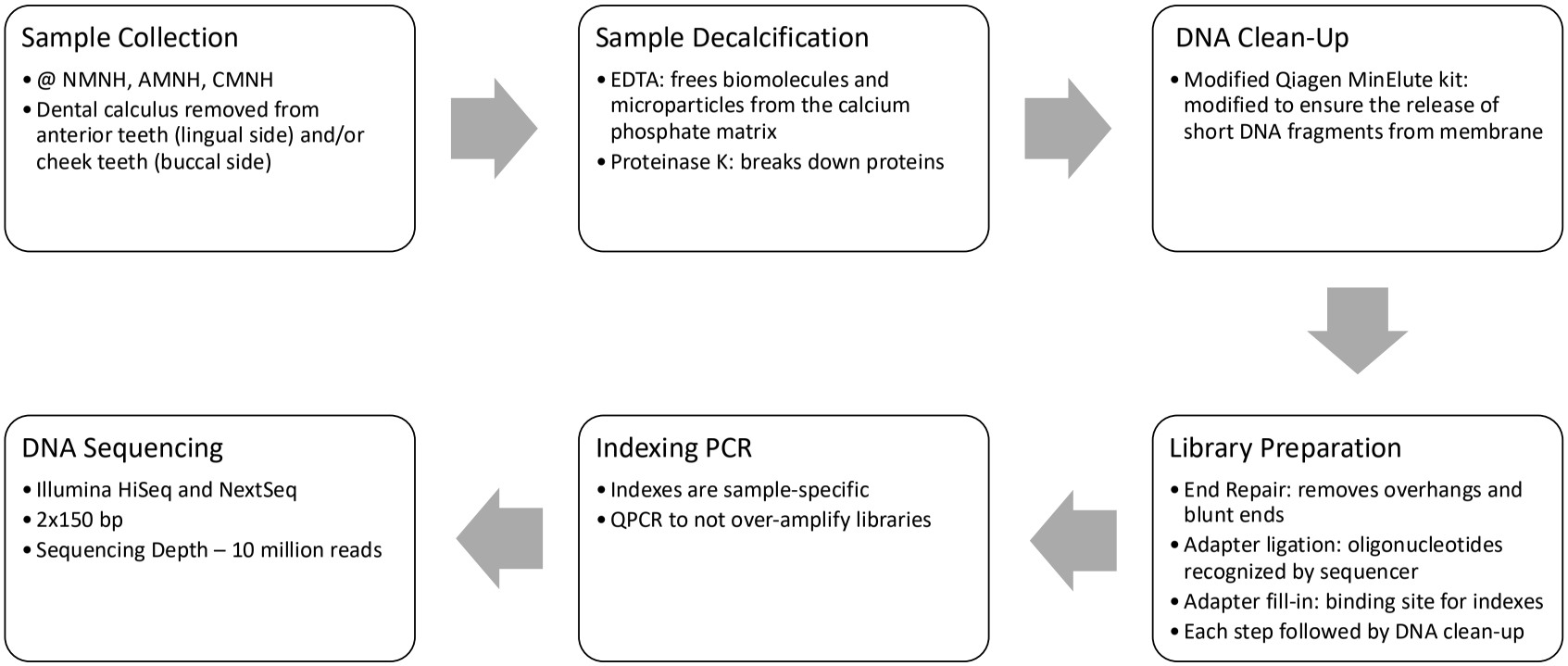
[a]only accounts for cavitated teeth
[b]also referred to as AMTL intensity
[c]used scoring system from Molnar 1971
[d]by individual

[e]crude prevalance of dental conditions
[f]total dentition not given

Museum Acc.	Museum	Collection	Race	Sex	Age	Year of Death	Cause of Death
HTH128	CMNH	Hamann-Todd	N	F	35	1918	Myocarditis
HTH1858	CMNH	Hamann-Todd	O	M	47	1929	Paralysis of the Insane (Syphilis)
HTH1875	CMNH	Hamann-Todd	O	M	43	1929	Paralysis of the Insane (Syphilis)
HTH215	CMNH	Hamann-Todd	W	M	28	1914	Pulmonary TB
HTH256	CMNH	Hamann-Todd	W	M	22	1914	Acute Cerebralspinal meningitis
HTH267	CMNH	Hamann-Todd	W	M	25	1914	Pulmonary TB
HTH399	CMNH	Hamann-Todd	N	M	70	1916	Pnuemonia
HTH475	CMNH	Hamann-Todd	N	M	45	1916	Pnuemonia
HOW225475	NMNH	Howard	N	M	Adult	NA	NA
HOW228039	NMNH	Howard	N	F	Adult	NA	NA
HOW248337	NMNH	Howard	N	M	Adult	NA	NA
HOW248635	NMNH	Howard	N	NA	Adult	NA	NA
HOW248636	NMNH	Howard	N	M	Adult	NA	NA
HOW248638	NMNH	Howard	N	M	Adult	NA	NA
HOW248639	NMNH	Howard	N	M	Adult	NA	NA
HOW248646	NMNH	Howard	N	M	Adult	NA	NA
HUNT182	NMNH	Huntington	W	M	Adult	NA	NA
HUNT289	NMNH	Huntington	W	F	Adult	NA	NA
HUNT396	NMNH	Huntington	W	F	Adult	NA	NA
HUNT227864	NMNH	Huntington	W	NA	Adult	NA	NA
HUNT317763	NMNH	Huntington	W	M	29	NA	NA
HUNT319344	NMNH	Huntington	W	M	32	NA	NA
HUNT319754	NMNH	Huntington	W	M	39	1895	PHTHISIS
HUNT319865	NMNH	Huntington	N	M	Adult	1900	PNEUMONIA
KOD364455	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364456	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364459	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364438	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364440	NMNH	Kodiak	O	M	Adult	NA	NA

KOD364441	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364443	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364445	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364449	NMNH	Kodiak	O	M	Adult	NA	NA
KOD364451	NMNH	Kodiak	O	M	Adult	NA	NA
MED98137_Q01	AMNH	Medical Collection	W	M	56	1946*	NA
MED98153_Q01	AMNH	Medical Collection	W	M	54	1946*	NA
MED98298_Q01	AMNH	Medical Collection	W	M	58	1946*	NA
MED98302_Q01	AMNH	Medical Collection	W	M	55	1947**	NA
MED98362_Q01	AMNH	Medical Collection	N	M	38	1947**	NA
MED98373_Q01	AMNH	Medical Collection	N	F	65	NA	NA
MED98376_Q01	AMNH	Medical Collection	W	M	40	NA	NA
MED9895_Q01	AMNH	Medical Collection	N	M	56	1945*	NA
HOW350606	NMNH	Patuxent R, Md.	N	F	Adult	NA	NA
HOW350607	NMNH	Patuxent R, Md.	N	F	Adult	NA	NA

Supplementary Table 3.1: Sample Information for Chapter 3.
* indicates dates the skeletal remains were received and accessioned into the collection.
** indicates dates measurements were taken of the skeleton



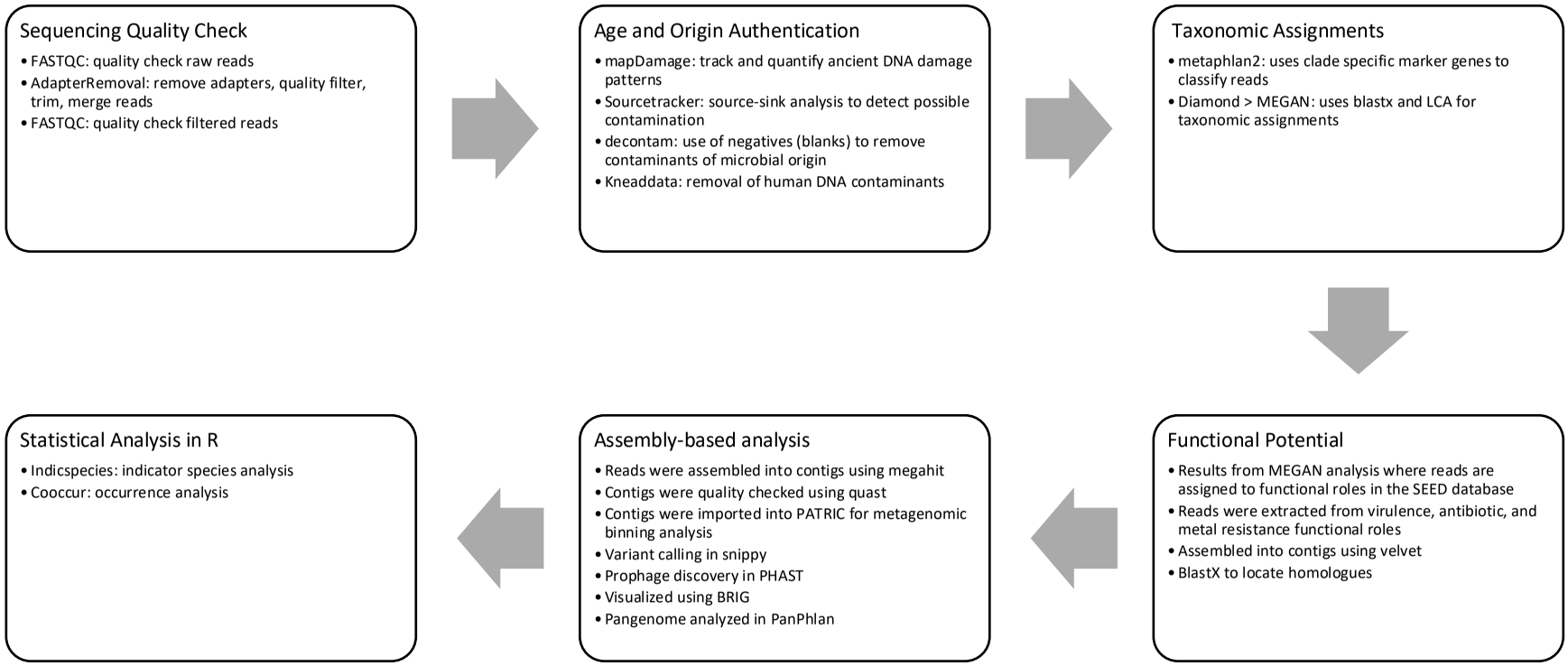
Supplementary Figure 3.2: Detailed Materials and Methods Flow Chart

Sample ID	Env	Reference
SRS013950	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014107	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014477	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014691	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015044	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015064	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015215	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015378	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015470	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS017227	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS018778	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS018975	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019029	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019129	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019591	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS063215	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)

Supplementary Table 3.3a: Human Microbiome Project samples downloaded from the HMP browser (https://www.hmpdacc.org/resources/data_browser.php). Env stands for microbial environment.

SRA #	Reference
SRX340010	Warinner et al. 2014
SRX340011	Warinner et al. 2014
SRX340012	Warinner et al. 2014
SRX340013	Warinner et al. 2014
SRX340014	Warinner et al. 2014
SRX340015	Warinner et al. 2014
SRX340016	Warinner et al. 2014
SRX340017	Warinner et al. 2014
SRX340018	Warinner et al. 2014
SRX2839918	Willman et al. 2018
SRX2839919	Willman et al. 2018
SRX2839921	Willman et al. 2018
SRX2839922	Willman et al. 2018
SRX2839923	Willman et al. 2018
SRX2839924	Willman et al. 2018
SRX2839925	Willman et al. 2018
SRX2839926	Willman et al. 2018
SRX2839927	Willman et al. 2018
SRX3745293	Willman et al. 2018
SRX3745294	Willman et al. 2018
SRX3745295	Willman et al. 2018

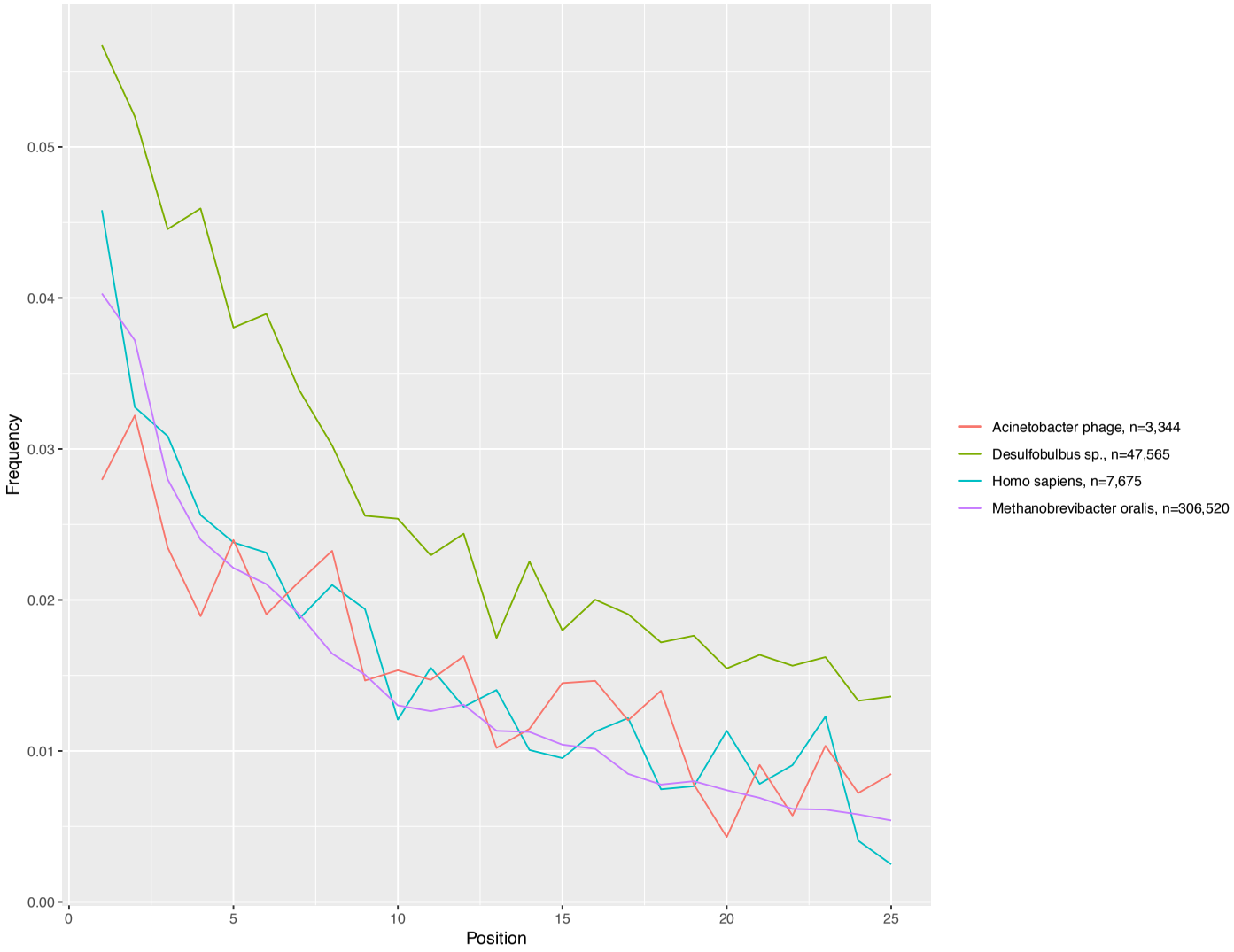
Supplementary Table 3.4: Medieval[9] and historical[10] dental calculus metagenomes downloaded for comparative indicator analysis in Chapter 4



Supplementary Figure 3.5: Detailed Bioinformatics Flow Chart

Taxa	Reference Number
Homo sapiens nuclear genome	hg19
Homo sapiens mitochondrial genome	NC_0129201
Acinetobacter baumannii	ASM74664v1
Bacteriodetes oral taxon 272	NZ_AUTU01000030.1
Bulkholderia cepacia	NZ_CP012981.1
Desulfobulbus oralis	CP021255.1
Eubacterium saphenum	NZ_ACON00000000.1
Fretibacterium fastidiosum	FP929056.1
Fusobacterium nucleatum	NC_003454.1
Methanobrevibacter oralis	GCA_001639275.1
Methanobrecibacter smithii	NC_009515.1
Olsenella uli	NC_014363.1
Porphyromonas gingivalis	NC_010729.1
Pseudopropionbacterium propionicum	NC_018142.1
Pseudoramibacter alactolyticus	NZ_GL622359.1
Streptococcus mutans	NC_004350.2
Trepomena denticola	NC_002967.9
Tannerella forsythia	NC_016610.1
Treponema maltophilum	NZ_KE332518.1
Treponema socranskii	NZ_KE332515.1
Acinetobacter baumannii phage YMC/09/02/B1251 ABA BP	NC_019541.1
Bacillus virus G	NC_023719.1

Supplementary Table 3.7: List of species used in mapdamage analysis



Supplementary Figure 3.8: C to T Substitutions across all domains of life. N indicates the number of reads mapped to each species listed by sample.

Supplementary Information 3.9: Further Details of metaphlan2 Analysis for Chapter 3

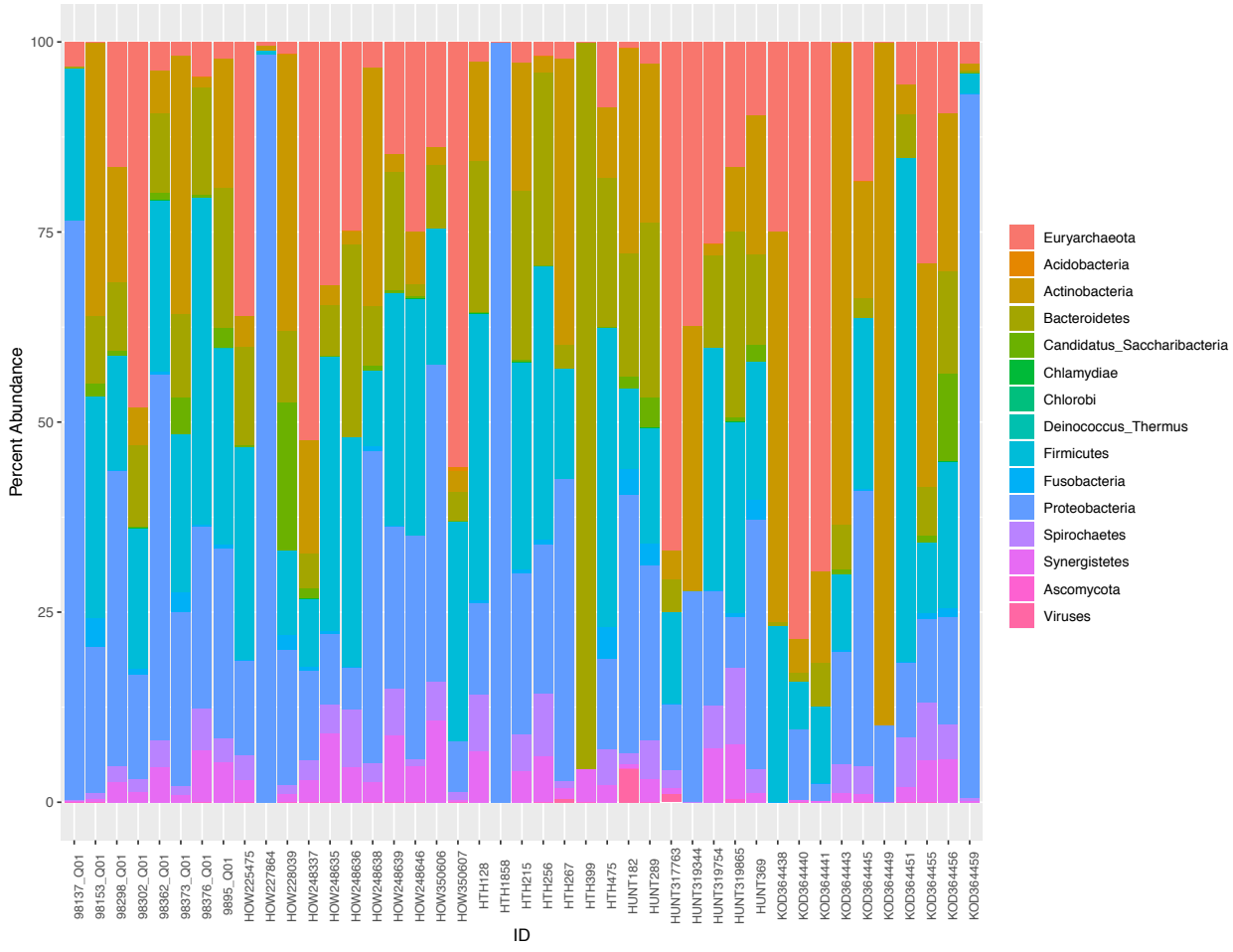
Historical calculus metagenome analysis

A total of 675,937 reads were removed from samples, leaving 225,096,484 for read- and assembly-based analysis in human contaminant analysis. Read-based analyses conducted in metaphlan2[1]. Prior to making taxonomic assignments, human DNA was removed using kneaddata (https://bitbucket.org/biobakery/kneaddata/wiki/Home). We observed a total of 856 taxonomic categories using metaphlan2[1]. The metaphlan2 database was published in 2015 and is comprised of around one million clade-specific marker genes for prokaryotic, viral, and microbial eukaryotic communities[1]. Metaphlan2 was used in this analysis infers both the presence and relative abundance of species, subspecies, and strain-level marker genes from around 7,500 taxa[1].

Archaea accounted for an average of 19% of the reads per sample, ranging from 0-78.49% of characterized microbial communities in historic calculus. Bacteria represented on average 60% of the classified reads, ranging from 13-100% of samples. Members of Eukaryota were detected in two samples: HOW319865 and HUNT182. Both microbial eukaryotes in these samples belonged in the family Aspergillaceae, *Penicillium chrysogenum* (0.417% in HOW319865) and an unclassified member of Aspergillaceae (0.2271% in HUNT182). Viruses were also detected in historical calculus samples using metaphlan2. These includes *Alternanthera yellow vein betasatellite* (1.451% in HUNT317763), *Sauropus leaf curl disease associated DNA beta* strain PRJNA17633 (0.525% in HTH267), *Ageratum yellow vein Singapore alphasatellite* strain PRJNA14232 (4.45% in HUNT182), *Salmonella phage HK620* strain PRJNA14115 (0.24% in HUNT182), an unclassified strain of

Human adenovirus B (.003% in HOW248337), and an unclassified species of *T5likevirus* (4.09% in HUNT182).

Five phyla accounted for a majority of reads: Proteobacteria (average 25.6% per sample), Firmicutes (average 20.12% per sample), Euryarchaeota (average 17.56% per sample), Actinobacteria (average 16.42% per sample), and Bacteriodetes (average 11.6% per sample) as shown in Figure 1. Other prokaryotic phyla such as Spirochaetes, Synergistetes, Candidatus Saccharibacteria, Fusobacteria, Acidobacteria, Chlorobi, Chlamydiae, and Deinococcus Thermus account for the remaining phyla present in historical calculus samples. Additionally fungal phyla Ascomycota and viruses were detected as described previously.



Of the 865 taxonomic categories, 260 species were characterized in our samples and DNA processing blanks. We included 12 negatives during DNA extraction, library preparation, and QPCR analysis. Eight negatives were comprised of 100% unclassified taxa. Three of our extraction negatives were comprised of 3 species: *Propionibacterium acnes*, an unclassified species within *Granulicella*, and an unclassified species within *Burkholderia*. We detected 2 microbial species in one library negative: an unclassified species within *Escherichia* and Enterobacteria phage lambda.

In metaphlan2 analysis, 39/43 samples were dominated by an unclassified species of *Methanobrevibacter* accounting for 0-76.52% of species per sample. Other prominent species in historical calculus samples included *Eubacterium saphenum*, *Desulfobulbus sp oral taxon 041*, *Pseudoramibacter alactolyticus*, and *Bacteriodetes bacterium oral taxon 272*. “Red complex” pathogens, *Tannerella forsythia* and *Treponema denticola*, were also present. The five samples that lacked *Methanobrevibacter* contributions were KOD364443, KOD364449, HTH1858, and HTH399. We found *Actinomyces massiliensis* represented 58.36% and 89.8% of KOD364443 and KOD364449 respectively. HTH1858 in metaphlan2 analysis was comprised of 100% *Cronobacter sakazakii*. Lastly, HTH399 had two prominent species, *Bacteriodetes bacterium oral taxon 272* (95.58%) and *Pyramidobacter piscolens* (4.41%).

Metaphlan2 taxonomic classifications for all samples and negatives (blanks) is available in the supplementary spreadsheet document (Supplementary Table 3.7)

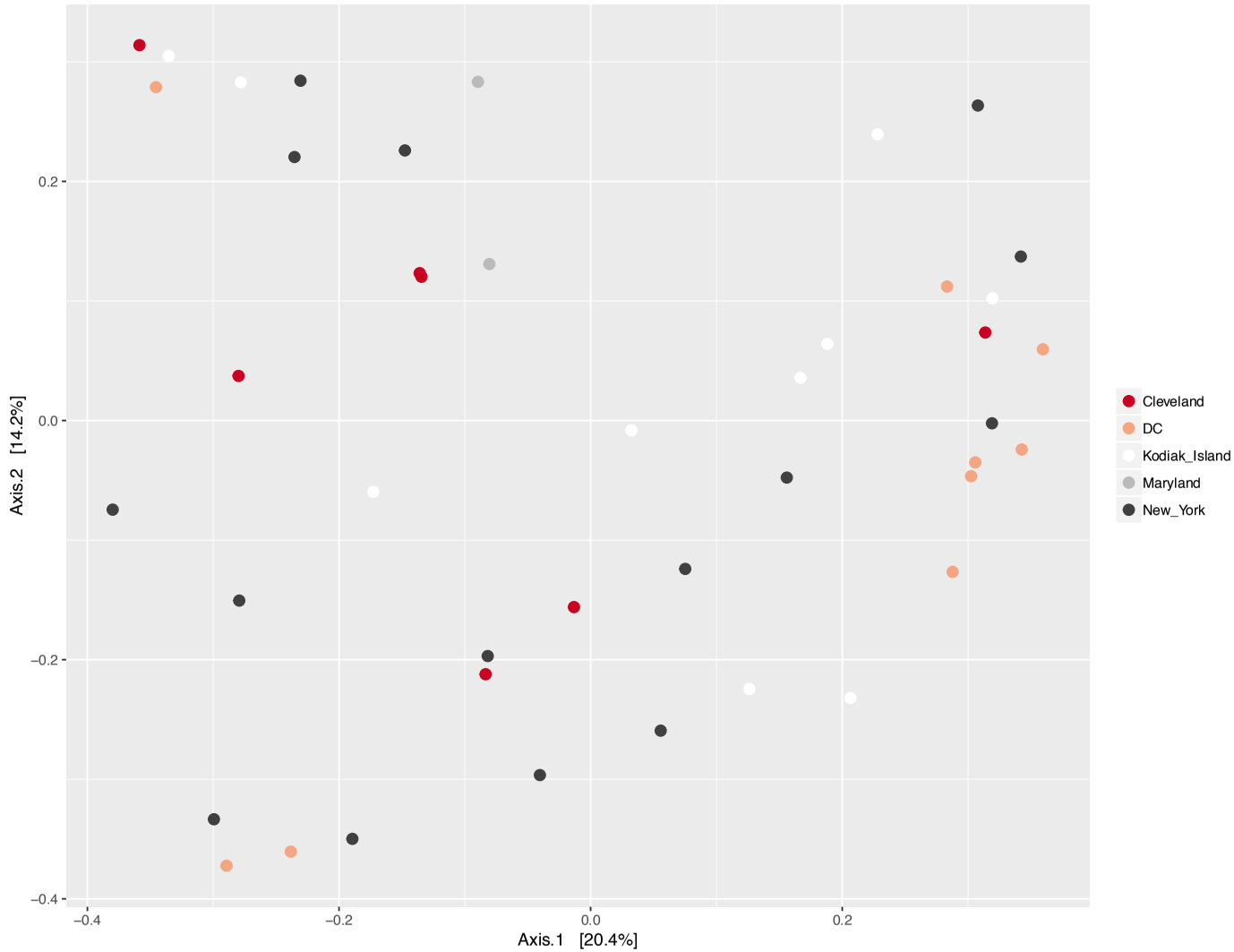
Indicator Species Analysis

To detect indicator species between historic-era calculus and modern plaque microbiota, supra- and subgingival plaque samples were also analyzed in metaphlan2 [2]. Out of 260 detected species, 88 species were found to be indicator species distinguish historical calculus and modern supragingival and subgingival plaque microbial communities. We detected 6 species to be significantly associated with historical calculus metagenomes: *Methanobrevibacter* unclassified, *Desulfobulbus sp. oral taxon 041*, *Eubacterium saphenum*, *Bacteriodetes bacterium oral taxon 272*, *Pseudoramibacer alactolyticus,* and *Olsenella uli* (all with p-values = 0.005). *Desulfobulbus* sp.*,* *Olsenella* sp., and *Methanobrevibacter* sp., and *Pseudoramibacer alactolyticus* overlap with findings from MEGAN, providing additional support to these taxa being associated with historical calculus microbiota across two different platforms.

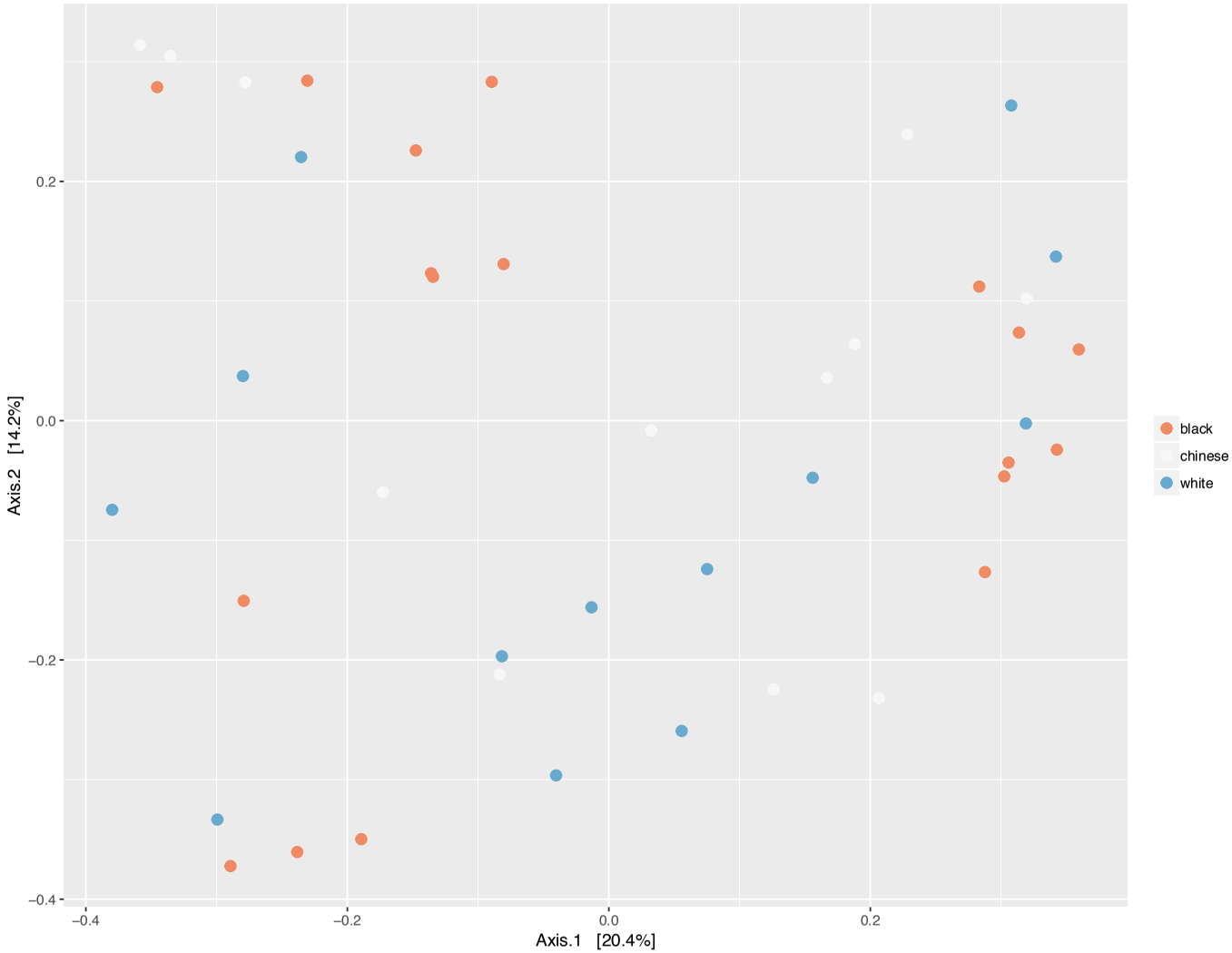
We found that 56 were significantly associated with both supragingival and subgingival plaque. The species identified as indicator species in HMP supragingival and subgingival plaque included *Streptococcus* sp., *Neiserria* sp., *Corynebacterium* sp., *Capnocytophaga* sp., and *Actinomyces* sp.. There were several species shared between metaphlan2 and MEGAN indicator species analysis. These were *Haemophilus parainfluenzae*, *Corynebacterium durum*, *Veillonella parvula*, *Porphyromonas cataoniae*, *Dialister invisus*, *Gemella morbillorum*.

References Cited

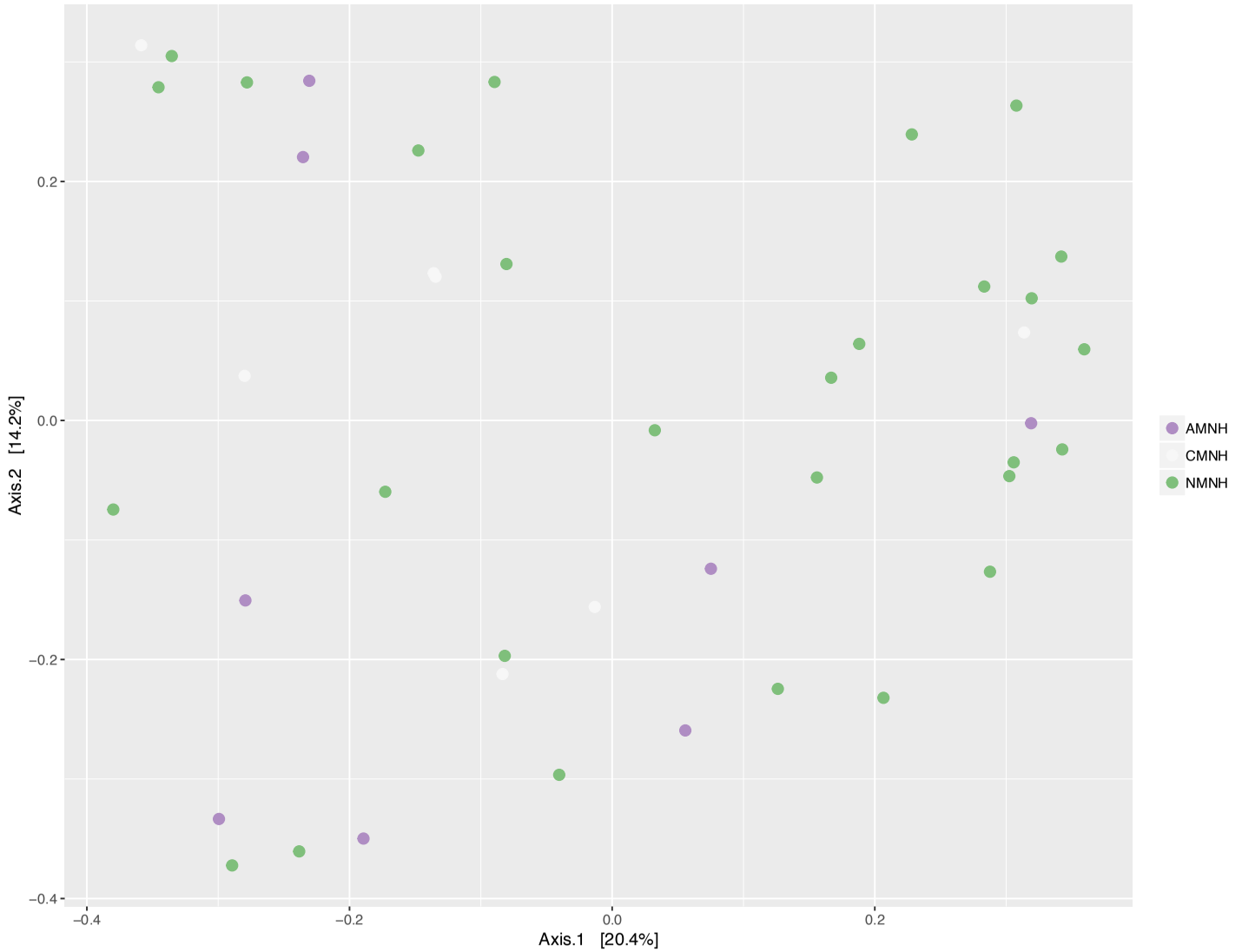
1. Truong, D.T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. **12**, 902 (2015).
2. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. **486**, 207 (2012).



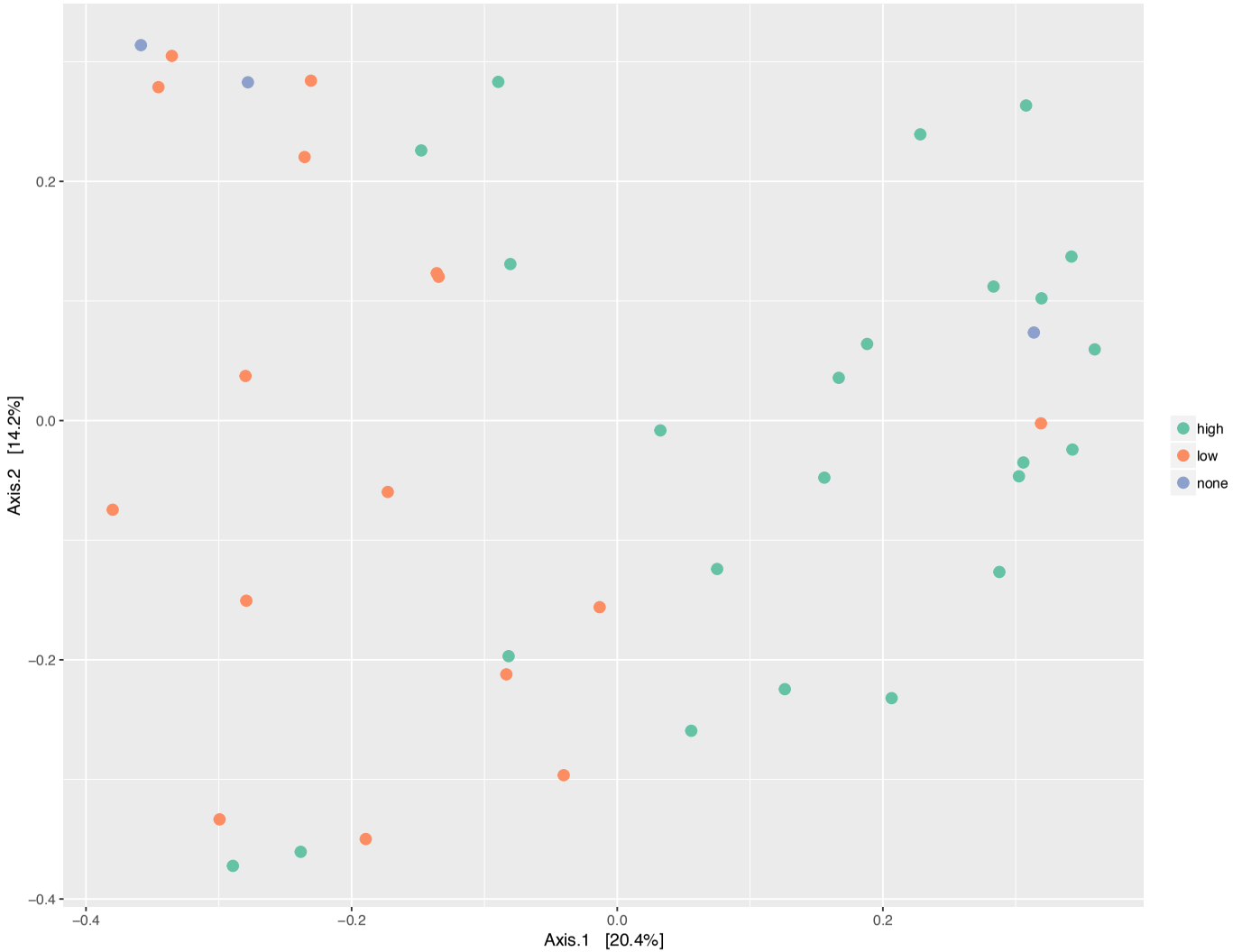
Supplementary Figure 3.15: PCoA plot constructed using Bray-Curtis distances. Here we show the beta diversity of dissertation samples by geographical origin.



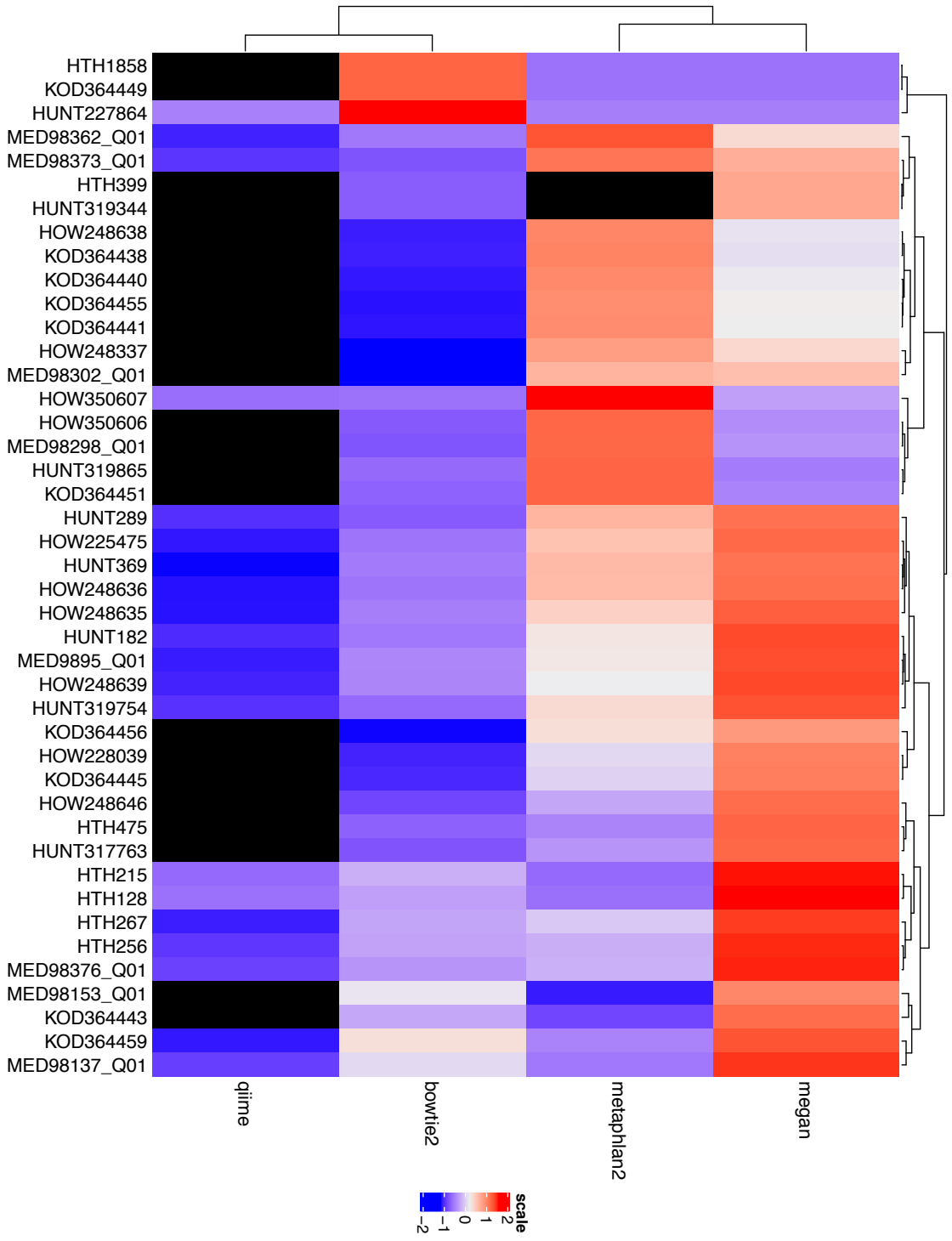
Supplementary Figure 3.16: PCoA plot constructed using Bray-Curtis distances. Here we show the beta diversity of dissertation samples by race.



Supplementary Figure 3.17: PCoA plot constructed using Bray-Curtis distances. Here we show the beta diversity of dissertation samples by museum collection.



Supplementary Figure 3.18: PCoA plot constructed using Bray-Curtis distances. Here we show the beta diversity of dissertation samples by amounts of archaea present in each sample assigned by MEGAN. High is above 5%, low is below 5%, and none is no archaea was present within the sample.


HTH1858
KOD364449
HUNT227864
MED98362_Q01
MED98373_Q01
HTH399
HUNT319344
HOW248638
KOD364438
KOD364440
KOD364455
KOD364441
HOW248337
MED98302_Q01
HOW350607
HOW350606
MED98298_Q01
HUNT319865
KOD364451
HUNT289
HOW225475
HUNT369
HOW248636
HOW248635
HUNT182
MED9895_Q01
HOW248639
HUNT319754
KOD364456
HOW228039
KOD364445
HOW248646
HTH475
HUNT317763
HTH215
HTH128
HTH267
HTH256
MED98376_Q01
MED98153_Q01
KOD364443
KOD364459
MED98137_Q01
qiime
bowtie2
metaphlan2
megan
scale
−2
−1
0
1
2

Supplementary Figure 3.19: Heatmap displaying relative abundances of *Methanobrevibacter oralis* assigned by megan (using LCA), metaphlan2 (uses clade-specific markers), and bowtie2 (each sample mapped to the PATRIC reference genome). For QIIME analysis, this heatmap displays 16S hits to the genus *Methanobrevibacter*. Black denotes two samples that were 100% unclassified in metaphlan2 analysis, HTH399 and HUNT319344 and samples with insufficient amount (>2.5k) of 16S data for analysis in QIIME.

Supplementary Information 3.17: Functional Annotations

Materials and Methods

Following decontamination, reads were aligned using DIAMOND to bacterial and archaeal reference protein sequences and all viral proteins available on GenBank (downloaded 3/2019, 70,800,065 amino acid sequences)[1]. Both decontaminated sequences and resulting DIAMOND text files were used to create RMA files for analysis in MEGAN[2]. The SEED mapping files were used for functional and virulence assignments. Reads were extracted from SEED functional roles using the "Extract Reads" function in MEGAN, assembled into contigs using velvet[3], and uploaded to NCBI blastx to locate homologous protein sequences[4].

Results

A total of 364,431 reads from 43 samples (average 8,591 hits per sample) were organized into functional categories in the SEED database; 18,358 reads were not able to be classified. All major modes of metabolism were represented in the remaining reads. The plurality were associated with carbohydrate metabolism which accounted from on average 17.07% of SEED hits (N=63,029). Amino acid metabolism accounted for 13.75% of SEED hits (N=50,931), followed by protein metabolism (7.85%, N=28,402). Sulfur metabolism (N=3,322), iron acquisition and metabolism (N=2,818), nitrogen metabolism (N=1,929), and potassium metabolism (N=1,866) were also present in our samples.

We observed 206 functional genes and enzymes associated with virulence available in Supplementary Table 3.21. A total of 12,517 contigs were

assembled from 126,584 reads extracted from MEGAN's SEED window from virulence and antibiotic resistance hits. The genetic material associated with virulence accounted for on average 2.53% of the functional annotations made against the SEED database. All blastx results are available in Supplementary Tables 2-29. Blastx results detailed in this section have e-values < 1E-10.

Three *Mycobacterium* virulence operons were present in all samples. These operons are involved in DNA transcription, protein synthesis, and quinolinate biosynthesis. Reads assigned to the *Mycobacterium* virulence operon DNA transcription were predominant DNA-directed RNA polymerase subunit beta with assembled contigs containing homologues in *Klebsiella* sp., *Pseudomonas* sp., *Escherichia* sp., and *Candidatus Saccharibacteria* sp. *Mycobacterium* virulence operon associated with protein synthesis were identified as elongation factor Tu, elongation factor G, GTP-binding protein, and methylmalonyl-CoA epimerase. Homologues from this function were observed in *Olsenella* sp., *Candidatus Saccharibacteria* sp., *Actinomyces* sp., and *Bulkholderia* sp. Homologues for quinolinate biosynthesis were associated with carboxylating nicotinate-nucleotide diphosphorylase and nicotinate-nucleotide diphosphorylase. These homologues hit to *Acinetobacter baumanni*, *Olsenella sp. oral taxon 207, Neisseria sicca*, and *Lautropia mirabilis*.

Heat shock protein GroEL and GMP synthease (EC 6.3.5.2) was present in all but one sample . Contigs assembled from the heat shock protein subsystem hits included *Fusobacterium periodonticum*, *Cardiobacterium hominis*, *Selenomonas sp. oral taxon 920*, *Olsenella* sp, and *Streptococcus* sp. GMP synthase homologues were observed in *Streptococcus oralis*, *Peptostreptococcaceae bacterium oral taxon 113*, *Fretibacterium sp. H1220_COT-178*, *Lautropia mirabilis*, and *Ottowia sp. Marseille-P4747*.

Multidrug resistance efflux pumps were ubiquitous and present in all but one sample. MATE family efflux transporter homologues were recovered from *Campylobacter rectus*, *Peptoniphilus grossensis*, *Anaerolineaceae bacterium oral taxon 439*, *Methanobrevibacter sp. YE315*, *Streptococcus mitis*, *Catonella* sp., and *Eubacterium* sp. Additionally, genes encoding resistance to acriflavin, floroquinolines, and vancomycin were also present in historical calculus samples. Acriflavin resistance genes were homologous to efflux RND transporter permease subunits from *Bacteriodes oral taxon 274*, *Tannerella forsythia*, *Stenotrophomonas maltophila*, *Treponema socranskii*, and *Capnocytophaga* sp. Contigs are also homologues of acriflavin resistance protein B from *Hyphomicrobium* sp., *Rhodoplanes roseus*, and *Afipia sp. 6-7*. Assembled contigs hit to DNA gyrase subunits A and B. We identified gyrase A hits from *Candidatus Saccharibacteria* sp., *Lautropia* sp., and a *Desulfovibrionaceae* bacterium. Homologues from gyrase B were also identified as DNA topoisomerase subunit B. Homologues identified in this SEED subsystem were from *Bulkholderia* sp., *Actinomyces* sp., *Tannerella* sp., *Lautropia* sp., *Cornyebacterium* sp. Resistance to vancomycin were assigned to the vanR and vanW genes in SEED. Homologues to the vanR gene were detected in *Olsenella* sp. and *Actinomyces* sp. Proteins homologous to VanW were found in *Paenibacillus* sp. and *Candidatus Yanofskybacteria* sp.

We identified beta-lactamase (E.C. 3.5.2.6) in 23 individuals. Ultimately, twenty-one contigs hit to beta-lactamases. Class A beta-lactamases are homologous to proteins in *Burkholderia cenocepacia* and *Streptococcus* sp. Contigs hit to class D-beta lactamase from *Acinetobacter baumannii*. Three contigs within this subsystem also hit to serine hydrolases from *Acinetobacter nosomialis*, *Klebsiella aerogenes*, and *Clostridium amazonitimonense*.

We also recovered genes associated with metal resistance. Arsenic, copper, cobalt-zinc-cadmium, and zinc resistance genes were recovered in our samples. A single contig was homologues to the ArsA family ATPase in *Lautropia* sp. Contigs assembled from the copper resistance subsystem contained homologues to heavy metal translocating P-type ATPase from *Streptococcus intermedius JTH08* and *Olsenella sp. oral taxon 807*. We also observed homologues from copper translocating P-type ATPse from *Methanobrevibacter oralis*, *Brachybacterium nesterenkovii*, *Leptotrichia sp. oral taxon 879*, and *Clostridium* sp. Contigs assembled from the cobalt-zinc-cadmium resistance SEED are homologous to CusA/CzcA family heavy metal efflux RND transporter from *Methylobacterium currus*, *Enterococcus faecalis*, *Pseudomonas aeruginosa*, *Prevotella* sp, and *Lautropia* sp. Homologues to the zinc resistance response regulator sigma 54 were identified in *Stenotrophomonas maltophila*, *Escherichia coli*, *Desulfobulbus* oralis, and *Klebsiella pneumoniae*.

Phages, prophages, transposable elements, and plasmids accounted for on average 0.09% of functional annotation. A total of 12 contigs (assembled from 1,307 reads) were assigned to phage portal proteins within SEED. Phage portal proteins in our dataset contained homologues currently found in *Paenibacillus phage PG1*, *Lachnospiraceae bacterium AM25-17* and *FE2018*, and *Bacillus oleronius*. We observed 500 hits to integrase/recombinase genes from oral bacteria such as *Streptococcus* sp., *Treponema* sp., and *Eubacterium* sp., however all e-values were too high to fit criteria for inclusion.

Blastx results are available on Supplementary Tables 3.21-3.39.

Discussion

In addition to whole genomes, we reconstructed genes associated with virulence and antibiotic resistance. There is evidence that oral microbiota has more diverse mobile genetic elements to adapt to the changing environment within the human mouth [5,6]. The oral cavity is a dynamic microenvironment with frequent fluctuations in temperature, pH, and oxygen. We found homologues of genes associated with Mycobacterium virulence, antibiotic resistance, metal resistance, and horizontal gene transfer.

Virulence factors are enriched in periodontal disease[7]. The oral cavity has a constant temperate of 35-36^{o}C and increases in periodontal pockets[8]. Heat shock proteins allow bacteria to adapt increases in temperature and other environmental stressors[9]. Temperature fluctuations also occur during the consumption of hot and cold food[10]. GroEL is upregulated in oral microbiota, such *Fusobacterium nucleatum* and *Porphyromonas gingivalis*, following heat stress [9,11]. In *Streptococcus mutans*, GroEL levels also rises following acid stress[10,12]. GroEL-like proteins in *A. actinomycetemicomitans* may play a role in disease initiation by increasing cell proliferation in low concentrations and has toxic effects on epithelial cells[13].

Metal resistance genes are among those overrepresented in diseased oral microbiota[7]. Mercury and cobalt-zinc-cadmium resistance were found to be enriched in periodontal samples[7]. Out of the 508 contigs assembled from the cobalt-zinc-cadmium SEED subsystem, we observed 29 hits to prokaryotic species, 18 of which were of high quality. Homologues belonged to inhabitants of the oral cavity, such as *Cardiobacterium hominis*, *Prevotella* sp., and *Tannerella* sp. Additionally, homologues of transients such as *Acinetobacter* and *Pseudomonas* sp. were detected in our calculus samples.

The role of antibiotic resistance in periodontal disease is unclear since they are present in healthy and diseased samples[7]. Homologues for resistance to acriflavin, floroquinolines, and vancomycin were detected in calculus samples in addition to the MATE family efflux transporter, which was present across a majority of samples. A majority of the individuals in our study died well before the first antibiotic, penicillin was introduced in 1942[14]. Resistance genes detected are not those commonly used to treat orofacial infections such as amoxicillin, metronidazole, erythromycin, clarithromycin, tetracycline, clindamycin, and fusidic acid[8]. This indicates the maintenance of these resistance genes within the oral cavity was long-established prior to the use of antibiotics as described in previous research[15].

References Cited

1. Buchfink, B., Xie, C. & Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59 (2015).
2. Huson, D.H., Auch, A.F., Qi, J. & Schuster, S.C. MEGAN analysis of metagenomic data. *Genome Res* **17**, 377-86 (2007).
3. Zerbino, D.R. Using the velvet de novo assembler for short-read sequencing technologies. *Current protocols in bioinformatics* **31**, 11.5. 1-11.5. 12 (2010).
4. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
5. Zhang, Q., Rho, M., Tang, H., Doak, T.G. & Ye, Y. CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes. *Genome biology* **14**, R40 (2013).
6. Roberts, A.P. & Kreth, J. The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome. *Frontiers in cellular and infection microbiology* **4**, 124 (2014).
7. Liu, B. *et al.* Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS One* **7**, e37919 (2012).
8. Marsh, P.D., Martin, M.V., Lewis, M.A. & Williams, D. *Oral Microbiology*, (Elsevier health sciences, 2009).
9. Gophna, U. & Ron, E.Z. Virulence and the heat shock response. *International journal of medical microbiology* **292**, 453-461 (2003).
10. Goulhen, F., Grenier, D. & Mayrand, D. Oral microbial heat-shock proteins and their potential contributions to infections. *Critical Reviews in Oral Biology & Medicine* **14**, 399-412 (2003).
11. Mayhew, M. *et al.* Protein folding in the central cavity of the GroEL–GroES chaperonin complex. *Nature* **379**, 420 (1996).
12. Hrimech, M., Mayrand, D., Grenier, D. & Trahan, L. Xylitol disturbs protein synthesis, including the expression of HSP-70 and HSP-60, in Streptococcus mutans. *Oral microbiology and immunology* **15**, 249-257 (2000).
13. Zhang, L. *et al.* Bacterial heat shock protein-60 increases epithelial cell proliferation through the ERK1/2 MAP kinases. *Experimental cell research* **266**, 11-20 (2001).
14. Grossman, C.M. The First Use of Penicillin in the United StatesThe First Use of Penicillin in the United States. *Annals of Internal Medicine* **149**, 135-136 (2008).
15. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nat Genet* **46**, 336-44 (2014).

Supplementary Information 3.41: Additional genomic analysis of reconstructed on *A. baumannii* and *S. maltophila* genomes

Two high quality bins were constructed in PATRIC from Proteobacteria-dominant calculus samples, *Acinetobacter baumannii* and *Stenotrophomonas maltophila*. We reconstructed a 95.3% complete genome (mean coverage 7.84) of *Acinetobacter baumanii* in MED98137_Q01. The GC content for the *A. baumannii* strain is 39.2% and is 3.4mbp long. It is comprised of 3832 coding regions and has 39 tRNAs. The AdaBoost classifier predicated a carbapenem susceptible resistant phenotype for this strain (accuracy 0.936, F1 score: 0.942, AUC:0.957). Specialty genes including virulence, antibiotic resistance, and drug targets were also identified. This includes 13 virulence factors encoding for biofilm synthesis and outer membrane secretin, catalyse KaTE, and ATP-dependent Clp protease. We observed a total of 40,079 variants in the historic-era strain: 6,327 complex variants, 179 deletions, 142 insertions, and 33,431 snps. Two prophage regions were identified, one 11.1kb region of an incomplete *Acinetobacter phage Bphi B1251* (NC_019541) and one 34.1kb region of a intact *Salmonella phage SSU5* (NC_018843.1).

Pangenome analysis of historic-era *A. baumannii* strain (median coverage 10) with 123 complete genomes available in GenBank revealed unique gene families only present in our reconstruction. In the reconstructed genome from MED98137_Q01, 3,471 gene families were recovered, out of an average of 3819 present in reference genomes. Reference genomes shared between 65.9-78.7% of gene families with AMNH98137_Q01. The reference sequence (GCF_001721705.1) with the highest percentage of similarity at 78.7% was recovered in 2010 from an Australian person with pneumonia.

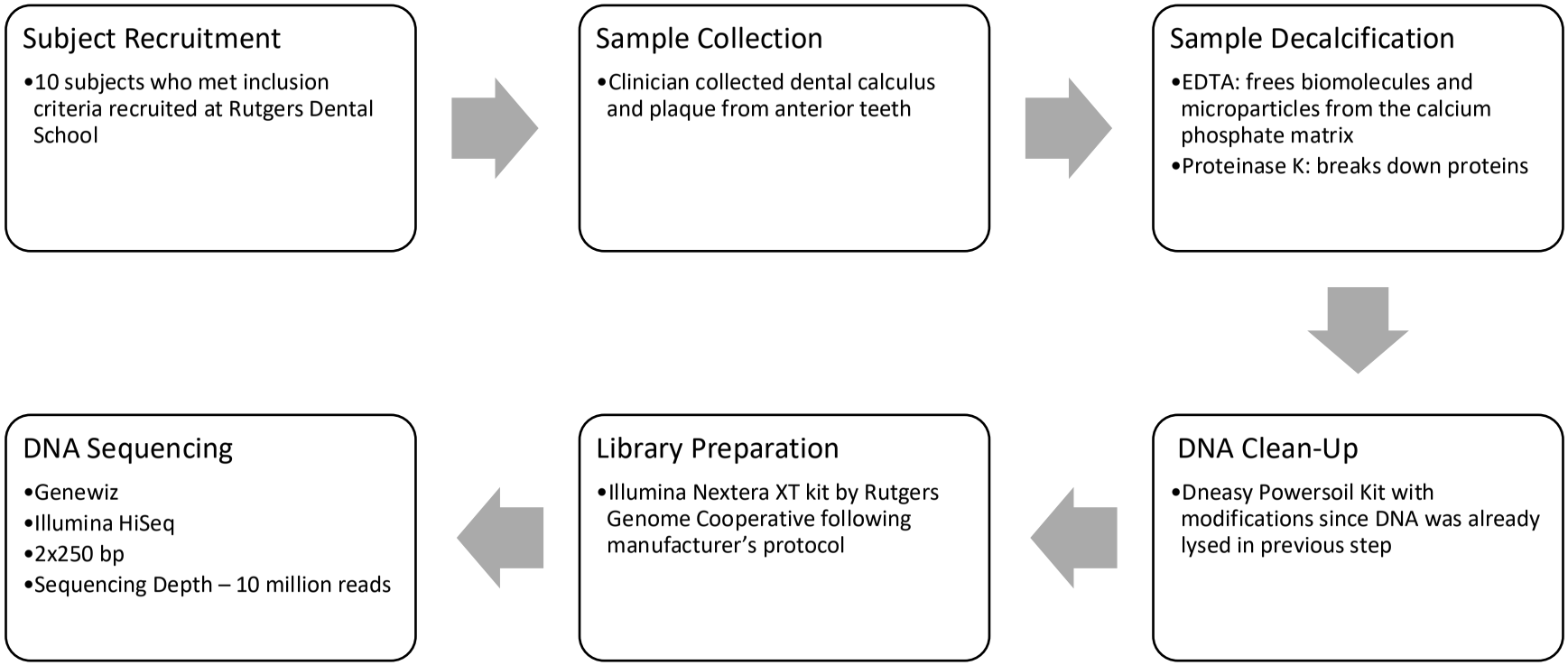
Stenotrophomonas maltophila

A 98.1% complete genome (mean coverage 16.94) of *Stenotrophomonas maltophila* was reconstructed from a single individual, KOD364459. The GC content for *S. maltophila* reconstruction is 66.35% and is 4.5mbp long. It is comprised of 4,815 coding regions and 57 tRNAs. Specialty genes including virulence, antibiotic resistance, and drug targets were identified. Virulence genes that encode elongation factor Tu and twitching motility protein PilG were found. Forty seven antibiotic resistance genes were found; these include multidrug efflux systems EmrAB-OMF and MdtAMC-TolC, isoleucyl-tRNA synthetases, and beta-lactamases. Finally, eight modern drug targets were also found in the reconstructed strain, these include acyl carrier protein, thiol peroxidase, and N-acetylmithine carbamoyltransferase. We observed a total of 40,042 variants in our reconstructed S. maltophila strain: 6,401 complex variants, 152 deletions, 156 insertions, and 33,333 snps. Two intact prophage regions were identified with the reconstructed strain: a 33.5kb region of *Stenotrophomonas phage Smp131* with 31 ORFs and 40.7kb region of *Mesorhizobium phagevB_MloP_LO5R7ANS*.

We conducted a pangenome analysis of a single *Stenotrophomonas maltophila* genome recovered from individual KOD364459 with 22 reference strains. A total of 4,164 gene families were recovered from our reconstruction and an average of 4,287 were shared by reference strains. Reference genomes shared between 6.2%-82.3% of gene families with our reconstructed strain. The reference sequences with the highest percentage of shared gene families (GCF_900475405.1; GCF_900475685.1) at 82.3% and 81.1% respectively was collected in 1957-1961 from cerebrospinal fluid and an oropharyngeal swab.

Supplementary Information 4.1: Modifications to DNeasy PowerSoil kit

We modified DNeasy PowerSoil protocol upon consultation from the help desk. Qiagen personnel suggested that we centrifuge the decalcification product at 13krpm for 1 minute. We then added ~1ml of the supernatant, following decalcification, to the a clean 2ml collection tube (included in the kit), and proceeded with the detailed protocol at step 10.



Supplementary Figure 4.2: Detailed Materials and Methods Flow Chart

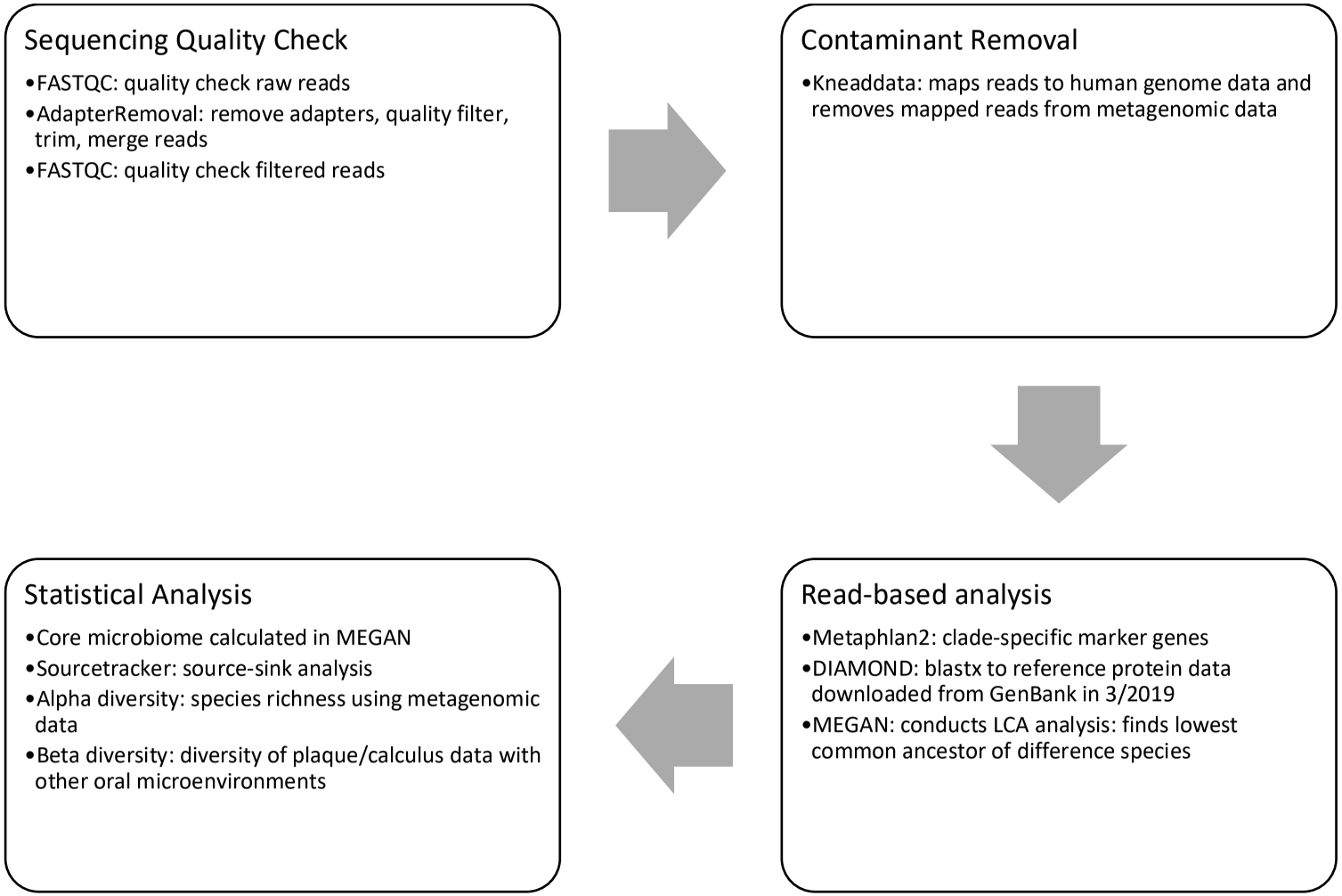
Sample ID	Env	Reference
SRS011105	anterior nares	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS011263	anterior nares	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS012291	anterior nares	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS012663	anterior nares	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013876	anterior nares	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS016503	buccal mucosa	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS023591	buccal mucosa	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS023837	buccal mucosa	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS024641	buccal mucosa	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS045049	buccal mucosa	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013946	keratinized gingiva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014473	keratinized gingiva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014687	keratinized gingiva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015060	keratinized gingiva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019025	keratinized gingiva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013947	palatine tonsils	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014474	palatine tonsils	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015061	palatine tonsils	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019026	palatine tonsils	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019126	palatine tonsils	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013942	saliva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014468	saliva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014692	saliva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015055	saliva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019120	saliva	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013950	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)

SRS014107	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014477	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014691	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015064	subPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015044	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015215	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015378	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015470	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS018778	supPlaque	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013948	throat	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014475	throat	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014689	throat	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015062	throat	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS019027	throat	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS011306	tongue dorsum	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS013234	tongue dorsum	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS014573	tongue dorsum	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015644	tongue dorsum	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)
SRS015893	tongue dorsum	The Human Microbiome Project Consortium 2012 (Huttenhower et al. 2012)

Supplementary Table 4.3: Human Microbiome Project samples downloaded from the HMP browser (https://www.hmpdacc.org/resources/data_browser.php). Env stands for microbial environment.

SRA #	Biosample #	Local environmental context	Collection Date	Geographic Location
SRS832559	SAMN03287614	a sample from a patient with chronic Periodontitis S01-01-D	2010	Los Angeles
SRS830523	AMN03287622	a sample from a patient with chronic Periodontitis S04-05-R	2010	Los Angeles
SRS830004	SAMN03287626	a sample from a patient with chronic Periodontitis S07-01-D	2010	Los Angeles
SRS833646	SAMN03287588	a sample from a patient with chronic Periodontitis S13-03-R	2010	Los Angeles
SRS840786	SAMN03287599	a sample from a patient with chronic Periodontitis S14-02-D	2010	Los Angeles
SRS833651	SAMN03287592	a sample from a patient with chronic Periodontitis S17-03-R	2011	Los Angeles
SRS833656	SAMN03287597	a sample from a patient with chronic Periodontitis S20-04-R	2010	Los Angeles
SRS841170	SAMN03287605	a sample from a patient with chronic Periodontitis S21-04-R	2011	Los Angeles
SRS841173	SAMN03287608	a sample from a patient with chronic Periodontitis S27-03-R	2011	Los Angeles
SRS841386	SAMN03287613	a sample from a patient with chronic Periodontitis S31-04-R	2010	Los Angeles

Supplementary Table 4.4: List of periodontal metagenomes downloaded in comparative analysis



Supplementary Figure 4.5: Detailed Bioinformatics Flow Chart