UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPACT OF HOST DIET AND LIFESTYLE ON ORAL MICROBIAL ECOLOGIES ACROSS TEMPORAL AND GEOGRAPHIC SCALES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By SARAH J. JOHNSON

Norman, Oklahoma

2025

IMPACT OF HOST DIET AND LIFESTYLE ON ORAL MICROBIAL ECOLOGIES ACROSS TEMPORAL AND GEOGRAPHIC SCALES

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ANTHROPOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Patrick Livingood, Chair

Dr. Tanvi P. Honap

Dr. Courtney Hofman

Dr. Jessica Cerezo-Roman

Dr. Brandi Bethke

Dr. Daniel Becker

ACKNOWLEDGEMENTS

Thank you to all who supported me through this endeavor. I want to especially thank Dr. Cecil Lewis, my former advisor, for giving me the chance to demonstrate my abilities and grow as a scholar. His persistent support and patience allowed me to develop strength and confidence in myself. His dedication to including me in project development and collaborations has broadened my academic circle and is the basis upon which I have built my dissertation. I also thank Dr. Tanvi Honap for her unyielding encouragement and reassurance. I would not be where I am today without her guidance and friendship. I thank Dr. Krithi Sankaranarayanan for his encouragement in my computational abilities and Dr. Courtney Hofman for continuously highlighting ways to stand out as an academic and pursue methods of self-improvement. I thank Dr. Jessica Cerezo-Roman, Dr. Daniel Becker, and Dr. Brandi Bethke for their expertise in planning the structure of my dissertation and in building the knowledge base I have accumulated through my time here. I am especially grateful to Dr. Patrick Livingood for assuming the role of chair and providing invaluable feedback and perspective this past year.

Thank you to everyone at LMAMR, past (Kristin, Jacob, Dave, Hande, Justin, Cara, Amy, and Shrushti) and present (Robin, Horvey, Britt, Isabella, Alice, Chris, Karissa, Lilly, and all others) for their camaraderie, friendship, and scholarship. I am also grateful for all the collaborations I have had the fortune of engaging in, not only for including me in research projects, but also for the friendships I have forged along the way.

I am eternally grateful for the support of my family and friends, without whom none of this would be possible. Thank you to my mother, Terry, for being my greatest supporter

and cheerleader. Thank you to Katie for encouraging me to always grow and to accept myself, flaws and all. Thank you, Clarke, for her joy, and Adam for his pragmatism. I thank my sisters and brother, Holly, Hanna, and Edinho, for always being there to both gas me up and bring me back down to earth. Thank you to my father, Ed, for the unique experiences that drove me towards anthropology and to Vanessa for her constant encouragement.

I owe my success to each person I have crossed paths with on this journey, and I look forward to the days to come.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Microbiomes are complex host-associated microbial networks that play important roles in maintaining systemic health. These microbial communities are influenced by host behavior and environmental factors, while host health can be reciprocally affected by shifts in microbial composition and functioning. This dissertation investigates how the oral microbiome is impacted by factors such as host behavior, diet, geography, and time. This work uses calcified dental plaque, or dental calculus, which forms on the surfaces of the teeth throughout an individual's life and traps microbes and their genetic information. Leveraging advances in shotgun metagenomic sequencing and ancient DNA techniques, oral microbiomes were recovered from a total of 91 individuals from diverse ancestral populations. Chapter one provides the framework of the dissertation and discusses the current state of the field of ancient metagenomics, including the findings from other studies on how oral microbiomes impact health and are shaped by human bioculture. Chapter two investigates how the adoption of maize as a dietary staple by ancestral Maya populations in Belize led to changes in their oral microbiome over a nearly 10,000-year time period. Here, it is revealed that an increase in the availability of fermentable carbohydrates was accompanied by an increased necessity for acid tolerance among oral microbes, leading to differential abundance of key microbes. Chapter three investigates whether intensive acorn consumption impacted the oral microbiome of ancient Muwekma Ohlone peoples from California, U.S. in comparison to those of other populations of varying subsistence patterns. This chapter also includes an analysis of metagenomically assembled genomes (MAGs) from ancient American oral microbiomes, providing a

snapshot of oral microbial strain diversity in pre-European contact populations. The functional potential of metagenomes from those with this unique diet revealed a composition similar to agricultural groups, but without enrichment of specific organisms that mark the agricultural diet. Chapter four presents methods for extracting and interpreting non-human herbivorous oral metagenomes by examining dental calculus from ancient and modern bison in Oklahoma. This is the first investigation into oral microbiomes from bison dental calculus, and among the first to explore modern non-human dental calculus. This approach not only revealed prokaryotic organisms that mark bison oral microbiomes but also exhibited ecologically significant eukaryotic plant and fungal genomes accessible through bison dental calculus. Each project employed molecular techniques designated for ancient and modern DNA extraction and shotgun library preparation. Additionally, innovative computational methods for taxonomic and functional assessment, phylogenetic analysis, and *de novo* metagenome assembly expanded the breadth of knowledge that could be obtained from metagenomic data. By examining the interplay between diet and oral microbial composition and function, this dissertation summarizes the mechanisms driving the existence and prominence of oral microbes, especially some associated with oral disease. By reconstructing metagenomically assembled genomes, we uncover hidden strain diversity irrespective of host or clinical relevance. Finally, by expanding into non-human hosts, we demonstrate the acquisition of environmental eukaryotic DNA from dental calculus of a keystone mammalian species.

Keywords: bioinformatics, ancient DNA, oral microbiome, microbial ecology, diet, oral disease, phylogenetics, *de novo* metagenome assembly

CHAPTER 1 – INTRODUCTION

1.1 Oral microbiomes and One Health

Biological communities have varying degrees of complexity but operate within concentric domains of connectivity (Levin, 1992). As such, human health is situated within a larger One Health framework subject to connectivity within social structures and with domesticated and wild animals, in addition to being impacted by environmental pressures and resource availability (Steffens & Finnis, 2022; Wolf, 2015). These patterns of connectivity have impacted human populations since antiquity; thus, assessing human health over a temporal scale requires ecological and social context. Ecological interactions do not exclusively derive from external forces, but are sourced from, and reciprocally impact, host-associated microscopic networks, or microbiomes.

This dissertation explores life at a microscopic scale within a larger ecological context. Particularly, this work investigates the oral microbiome sourced from a type of archaeological specimen that can withstand taphonomic processes relatively well and preserve its biological constituents through deep time scales: calcified dental plaque, or dental calculus (Adler et al., 2013; Warinner et al., 2014). Calculus is a treasure trove of DNA from oral microbial, respiratory pathogen, host, and dietary components (Ziesemer et al., 2015). Oral microbiomes have evolved in concert with their respective hosts and are directly impacted by the interactions of their hosts with their environments (Brealey et al., 2020; Fellows Yates (B) et al., 2021). Host physiology is often robust to environmental variation through phenotypic plasticity, allowing for adjustments through temporary

periods of hardship (Kolodny & Schulenburg, 2020). Host associated microbiomes can mediate this mechanism through their own adaptive responses, contributing to evolutionary success of the host. Alternatively, host response can disrupt microbial function, leading to dysbiosis and disease (Leatherman & Goodman, 2020). This dialectical interplay between host and microbial biology and evolution has invited a multitude of theoretical perspectives on health, the body, and its position within the larger framework of evolution, history, and society (Formosinho et al., 2022; Gluckman et al., 2011; Leatherman & Goodman, 2020; Singer, 1996). The ecological one health model recognizes this blurred division between human, animal, and environmental health. This connectivity is further expanded in the rejection of a Cartesian divide between organism and environment and the critique of passive adaptation (Laland et al., 2016; Lewontin, 1983; Singer, 1996). Instead, it is argued that organisms select their environments and, under a 'niche construction' model, mold their surroundings to inadvertently shape the evolutionary trajectory of biotic and abiotic elements within the ecological niche (Laland et al., 2016; Lewontin, 1983; Odling-Smee et al., 1996). The host-microbiome relationship stretches the ontological definition of organism and environment (Formosinho et al., 2022).The 'situated biology' of bodies places the individual not only within an ecological niche, but accompanied by their life histories and those of their ancestors, as well as within the structures of society and ideology (Leatherman & Goodman, 2020; Lock, 2017). Microbiome evolutionary dynamics are likewise situated within the host sociocultural sphere, vulnerable to constraint by built environments and exposed to microbial diversity through host social behavior (Bosch et al., 2024; Sarkar et al., 2024). An 'environmentality'

perspective recognizes the arbitrary distinction between organism and environment (Formosinho et al., 2022). It directs the focus of the research question with respect to the different elements included in the biological network, determining the object of research and noting influential factors of the other elements (Formosinho et al., 2022). The dynamics of a microbiome are subject to the situated biology of the host, reciprocally shape the health and behavior of said host, and evolve within the confines of host social activity.

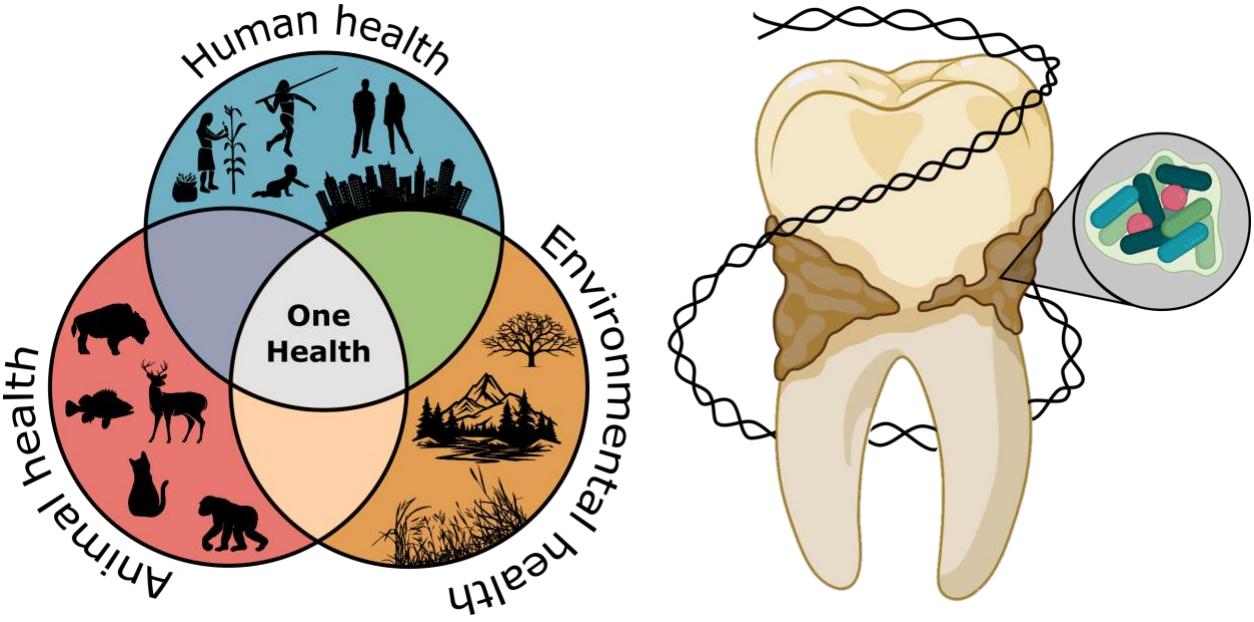


Figure 1.1 Visual summary.

Figure generated in Inkscape v1.0.2. Tooth and biofilm images acquired through BioRender (2024), and one health icons drawn or traced.

Dental calculus, which harbors microbial genetic material for millennia, can offer a window into the relationship between the host and their associated microbiota within different archaeologically informed contexts, situating a host's lifestyle within wider ecological networks.

Each project in this dissertation examines oral microbial ecologies, tying together the impact of host lifestyle and diet on microbial structure and function throughout history. The first project examines a compositional shift within the human oral microbiome in a population in the context of changing dietary practices. The second project investigates how the oral microbiome of a population with a unique dietary preference is differentiated from others and how it is impacted by major human biocultural events such as European contact. The final project of this dissertation explores oral microbiomes in a non-human host with a history deeply intertwined with human populations: the North American bison.

Much of this work involves the metagenomic reconstruction of oral microbiomes from human and non-human hosts to elucidate the evolutionary dynamics of individual microbes as well as the entire microbial communities. Despite investigating genomic diversity at a microbial level, our findings are interpreted within a much larger dynamic. The host populations studied here are separated by millennia, yet we find that individuals are connected through their oral microbial ecologies.

1.2 Microbial ecologies and oral disease

The oral microbiome is integral to both oral and systemic health outcomes. The progression towards oral disease is multifactorial, and the ecological plaque hypothesis suggests a correlation between host environment and disease (Marsh, 1994). This theory was further developed in 2012 with the keystone pathogen hypothesis, which implies that specific taxa, in low abundance and under certain environmental pressures, stimulate disruptive behavior of the normal flora (Hajishengallis et al., 2012). This dysbiosis mediates

the progression of oral disease with the demineralization of enamel and dentin, or dental caries (Scardina & Messina, 2012), and infection and inflammation of the surrounding gingiva, or periodontal disease (Abusleme et al., 2013). Additionally, dysbiotic oral microbial networks have been implicated in systemic disease beyond the oral cavity, including respiratory (Whiley & Hardie, 2015) and cardiovascular disease (Chhibber-Goel et al., 2016; Teles & Wang, 2011), cancer (Fan et al., 2018; Kato et al., 2016; Willis & Gabaldón, 2020), obesity (Sen et al., 2017), diabetes (Casarin et al., 2013), autoimmune disorders (Bruserud et al., 2018; Jensen et al., 2021; Li et al., 2020; Zhang et al., 2015), and Alzheimer's disease (Dominy et al., 2019; Miklossy, 2016). The mechanism by which host lifestyle and environment affects the composition and function of the oral microbiome is of great interest due to the widespread frequency of oral disease today. In the United States, 90% of the population suffers from dental caries and almost 50% of the population has experienced periodontal disease with much greater incidence among minority groups or those of lower socioeconomic status (National Institutes of Health, 2021).

The oral microbiome forms in a succession of communal assemblies from early to late colonizer bacterial species (Costello et al., 2012). Microbial colonization, or succession, is contingent upon the available niche, impediments or advantages available for dispersal of colonizing microbes, the compatibility between the species and the environment, and the resilience through variation in community composition (Chen et al., 2024). Succession is constrained by the environment in which the host exists, the historical contingency among colonizing species, the stochastic nature of available species reservoirs, and limitations to species dispersal (Costello et al., 2012). Within a

single mouth, there are variations in surface material, different levels of pH and exposure to oxygen, and different access to bulk fluid in the form of saliva (Socransky & Haffajee, 2005). Different mouths exist in different environments and belong to different host species, further diversifying the available ecosystems for oral microorganisms. Oral ecology is affected by dietary components, the level of food processing prior to consumption, and dental hygiene practices (Sedghi et al., 2021; Socransky & Haffajee, 2005).

In plaque and salivary studies, diet is shown to have a significant impact on the structure and function of oral microbiomes (Cordain et al., 2005; Halvorsrud et al., 2019; Millen et al., 2022; Nasidze et al., 2011). Millen and colleagues found that overall, carbohydrate consumption had an association with subgingival microbial composition in a cohort of post-menopausal women (Millen et al., 2022). They, and others, have also noted that the type of carbohydrates and other macronutrients consumed have differential effects on bacterial diversity and abundance (Halvorsrud et al., 2019; Kato et al., 2017; Millen et al., 2022). When an available niche is populated by a diverse array of microorganisms, or in other words there is a high degree of alpha diversity, the host is more capable of adapting to different environmental situations. High alpha diversity is thus associated with metabolic plasticity and is a marker of microbiome health (Sonnenburg & Sonnenburg, 2014). Plaque and salivary studies have shown that differential consumption of macronutrients can affect the composition of microbial communities, indicating that diet and lifestyle can distinguish microbial communities between groups (Belstrøm et al., 2014; Lommi et al., 2022). Consumption of processed starches, or rapidly digestible

starch, increases the risk of caries. This is due to the acidogenic nature of some opportunistic pathogens, such as *Streptococcus mutans*, decreasing pH and demineralizing enamel (Halvorsrud et al., 2019; Lommi et al., 2022). Slowly digestible starch and high fiber from whole grain foods has the opposite effect of mitigating the progression towards caries or periodontal disease (Nielsen et al., 2014). It is noted, however, that salivary nutrient and subsequent microbial composition shifts throughout the day, increasing the potential for variability in outcomes (Belstrøm et al., 2014; Kato et al., 2017; Lommi et al., 2022).

Oral disease is more pronounced in agricultural and industrialized societies. For example, when comparing salivary microbiomes from either modern hunters and gatherers (HG) in Uganda or the Philippines to neighboring farming (TF) and industrial groups, there was a prevalence of microorganisms associated with disease in the HG group, but there was a lower instance of oral disease relative to the TF and industrial groups (Lassalle et al., 2018; Nasidze et al., 2011). There is, however, greater alpha diversity in the HG groups compared to industrialized populations, with farming groups occupying intermediate diversity values (Lassalle et al., 2018; Nasidze et al., 2011). This supports the claim that a diverse system protects against dysbiosis from opportunistic pathogens and suggests that these taxa are commensal unless exposed to an environment that is preferential for pathogenicity functions. There has been a progressive decrease in alpha diversity across lifestyle transitions, such as the adoption of agriculture and progression towards industrialization (Sonnenburg & Sonnenburg, 2014). Additionally, the relative abundances

and prevalence of opportunistic pathogens has increased, along with the frequency of oral disease.

1.3 Ancient oral microbiomes

The calcified matrix of dental calculus harbors a high microbial biomass and has proven to protect DNA from taphonomic processes (Mann et al., 2018). Key characteristics of ancient DNA (aDNA) include short fragment lengths, lesions preventing DNA replication, and lesions causing incorrect base integration (Dabney (B) et al., 2013). The fragment breakage is due to the loss of a purine base, such as adenine (A) or guanine (G), followed by beta elimination, or where a bond from the DNA phosphate backbone is stripped by the depurinated sugar (Dabney (B) et al., 2013). These short fragments are overrepresented by A and G at the terminal nucleotides. The damage that is most often used in aDNA authentication is miscoding lesions, where an amine group is lost by cytosine (C), resulting in uracil (U) (Dabney (B) et al., 2013; Rohland et al., 2015). During replication, DNA polymerases read the U as a thymine (T) and place a complementary A, resulting in a high frequency of C to T and G to A transitions at terminal nucleotides. Although this damage is evident on DNA contained within dental calculus, the severity is lower compared to more exposed sources of microbial DNA, such as bone and dentin (Mann et al., 2018).

In a location directly in contact with dietary substrates, dental calculus has captured the attention of molecular anthropologists as a way to assess the impact of diet on oral microbiomes in antiquity. Calculus accumulates over the course of an individual's life, trapping biomolecules from the host, oral microbes, and microremains, such as

phytoliths and starch granules (Mann et al., 2018; Modi et al., 2020, 2023; Quagliariello et al., 2022; Velsko et al., 2019, 2022). This wholistic accumulation differentiates calculus composition from that of plaque or salivary microbiomes and thus is used to make broad assumptions about population-level dietary practice (Gancz et al., 2023; Ottoni et al., 2021; Quagliariello et al., 2022; Velsko et al., 2019). For instance, the increase in reliance on starch-rich foods among early *Homo* was accompanied by an increase in copy number of the starch-degrading α-amylase enzyme (*AMY1*) and subsequent enrichment of *Streptococcus* species capable of utilizing host salivary amylase (Fellows Yates (B) et al., 2021). The adoption of agriculture during the Neolithic transition in southern Europe did not show such a drastic compositional shift, likely due to the gradual nature of the Neolithic transition (Ottoni et al., 2021; Quagliariello et al., 2022). However, modernization significantly altered the oral microbial landscape, both with the adoption of processed foods and refined sugars, and the use of antibiotics (Ottoni et al., 2021). In Industrialized London, oral communities dominated by *Streptococcus* species had functions related to low-fiber, high-carbohydrate diets and were accompanied by the periodontal pathogens, *Porphyromonas gingivalis* and *Tannerella forsythia* (Gancz et al., 2023). With the relative stability of oral microbial communities, discerning host diet from affiliated metagenomes recovered from dental calculus has many limitations.

Calculus from modern populations that practice toothbrushing is primarily composed of early colonizers, like facultatively anaerobic and saccharolytic microbes, whereas calculus that has accumulated over a greater period of time without regular disruption of plaque biofilms has a greater proportion obligate anaerobes and

methanogens (Jin & Yip, 2002; Sedghi et al., 2021; Velsko et al., 2019; White, 1997). The formulaic microbial accumulation during calculus formation differentiates this taxonomic composition from other salivary or plaque microbiomes, necessitating equivalent comparative data for ecological interpretations (Velsko et al., 2019). Furthermore, the microbial composition of calculus includes taxa that may have been encapsulated within the matrix at any point during the lifetime of the host, negating any possibility of reconstructing transient responses to host lifestyle patterns without sectioning the calculus (Warinner et al., 2015).

Direct detection of genetic material related to dietary constituents from ancient dental calculus has largely been unsuccessful. The proportion of genetic material attributed to bacteria, archaea, and the host vastly outweighs other eukaryotic DNA that may be attributed to diet-related sources (Mann et al., 2023; Warinner et al., 2014; Weyrich et al., 2017). Reference databases used for taxonomic classification are overrepresented by clinically relevant, mostly microbial, organisms and do not fully capture the available biodiversity (Brealey et al., 2020). There is also known contamination among some eukaryotic genomes that are present in reference databases, such as Illumina adapter sequences in the reference carp genome (https://grahametherington.blogspot.com/2014/09/why-you-should-qc-your-reads-and-your.html) (Mann et al., 2023) and apicoplast DNA from Sarcocystidae parasites found in bird and bat genomes (Orosz, 2015). Additionally, the frequency of DNA read mis-mapping draws uncertainty about positive identification (Brealey et al., 2020; Mann et al., 2023; Ottoni et al., 2021; Weyrich et al., 2017). To circumvent mis-mapping and false positives,

some studies have relied on additional, corroborating information, such as plant microremains (Modi et al., 2023; Quagliariello et al., 2022), plant pathogens that typically accompany those plants of dietary relevance (Weyrich et al., 2017), recorded evidence of food consumed (Gancz et al., 2023; Moraitou et al., 2022; Quagliariello et al., 2022), or they have isolated their conclusions of dietary constituents to fauna that would have been available to the studied individual at the time (Moraitou et al., 2022; Weyrich et al., 2017). Authenticating ancient DNA identified as dietary in origin on its classification alone requires a higher frequency of nucleotide transitions at the terminal ends of the DNA reads and a breadth of coverage across the reference outside of conserved regions that would suggest alignment is not due to mis-mapping (Green et al., 2010; Orlando et al., 2021). These measures typically require at least 500 reads to be assigned to the eukaryotic species (Mann et al., 2023). Given the very low proportion of eukaryotic DNA present in dental calculus, this necessitates deeper sequencing. There has not yet been an example of animal DNA identified in dental calculus, though some studies have isolated plant DNA with varying degrees of success in authentication (Brealey et al., 2020; Mann et al., 2023; Moraitou et al., 2022; Ottoni et al., 2021; Weyrich et al., 2017). Reconstructing diet from ancient oral metagenomes requires an understanding of nuanced taxonomic and functional dynamics that result from dietary constituents over large time scales (Mann et al., 2023).

1.4 Dissertation Structure

This dissertation investigates the composition and function of oral microbial ecologies in the context of host, lifestyle, diet, geography, and time. Through a study of individual microbial genomes and their evolutionary trajectories, this work aims to highlight hidden microbial diversity that has been under-explored either due to a paucity of ancient oral microbiome data from the Americas, or because attention rarely deviates from human-associated, clinically relevant taxa.

In chapter 2, reconstructed oral microbiomes from ancestral Maya individuals from Belize are presented to investigate the impact of an agricultural transition on microbial composition and function. These rock shelters, located in the Maya mountains, show consistent deposition of human remains over a period of 10,000 years which spans the adoption of maize as a dietary staple by these peoples. This project is the first to explore the impact of such a major dietary transition on the oral microbiome of a population from the Americas.

In chapter 3, I present an analysis of oral microbiomes from Ancestral hunter-gatherers from California spanning an approximately 3,900-year time period. While most of these individuals pre-date European Contact and had a unique high-starch diet of acorns and seeds, some individuals are Mission-era agriculturalists, which allowed us to study the impact of European Contact and the resulting shifts in dietary practices on these oral ecologies. Additionally, we leverage the use of *de novo* metagenomically assembled genomes from these data to identify hidden strain diversity among oral microbes and better elucidate the evolutionary mechanisms of abundant and historically relevant taxa.

Chapter 4 involves a proof-of-concept exploration of ancient and modern bison dental calculus to demonstrate the utility of ruminant dental calculus as a window into present and past ecological diversity. Here, we supply methods to extract DNA from herbivorous ruminant calculus, making this the first study to process metagenomes from bison dental calculus and among the first to do so with modern non-human calculus in general (Brealey et al., 2021).

Through this work, I present analyses elucidating how microbial dynamics are shaped by host activity. I highlight how oral microbial communities are impacted by biological mechanisms tied to nutrient availability, host culture and practices, and geography, and emphasize the unique evolutionary trajectories of individual microorganisms situated within the larger context of human activity. This research was funded by a grant from the National Science Foundation (NSF BCS-2045308). Most of the computing for this dissertation research was performed at the OU Supercomputing Center for Education and Research (OSCER) at the University of Oklahoma (OU). All lab work was performed at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR).

CHAPTER 2 – SHIFTING ORAL MICROBIAL ECOLOGIES ACCOMPANY A DIETARY TRANSITION IN ANCESTRAL MAYA IN BELIZE

2.1 Abstract

Oral disease is mediated by dysbiotic oral microbial communities, the structure of which is influenced by host environment. Agricultural transitions in antiquity are associated with a decline in oral health, though the direct interaction between dietary shifts and microbial dynamics lacks nuance. Here, we investigated how the adoption of maize as a dietary staple in the neotropics altered the structure and function of oral microbiomes. Dental calculus was collected from the skeletal remains of 22 Ancestral Maya individuals recovered from two rock shelters in the Maya Mountains of Belize, which were continuously inhabited over 10,000 years across a transition to the adoption of maize as a staple food source. Extracted DNA was partially uracil-DNA-glycosylase-treated and double-stranded DNA libraries were shotgun-sequenced with Illumina technology. Taxonomic and functional composition was assigned to metagenomes and phylogenetic placement was determined for relevant taxa. The adoption of maize as a dietary staple by the Ancestral Maya was accompanied by a significant decrease in the abundance and functional contributions of the Clostridiales bacteria, *Pseudoramibacter alactolyticus* and *Eubacterium minutum*, and was replaced by saccharolytic *Streptococcus* species. This transition was also accompanied by an increase in abundance and diversity of mechanisms of acid tolerance and resistance. Despite its decrease in abundance, the *P.*

alactolyticus strain found in Ancestral Maya dental calculus persist till today, despite the introduction of distinct strains from Europe and Africa following colonization. Alternatively, pre-colonial era American strains of the opportunistic pathogen *Tannerella forsythia* were replaced following colonization, highlighting the divergent evolutionary histories of oral microbes. This study marks the first investigation of ancient oral microbial composition in response to a dietary transition in the Americas and demonstrates how human activity in antiquity shaped the evolutionary trajectory of oral microbiomes.

2.2 Introduction

The transition from hunting and gathering to an agricultural lifestyle irreparably altered the ecological niches in which humans reside (Kennett et al., 2020; Prufer et al., 2021). Though this transition prompted increased fecundity and technological innovation, these were not the only products of this changing subsistence economy. The gradual shift from hunting and gathering to horticulture and agriculture is also associated with a significant decline in overall health and a proliferation of oral pathologies (Adler et al., 2013; Cohen & Armelagos, 1984; Mummert et al., 2011). Today, oral diseases affect 3.5 billion people worldwide (World Health Organization, 2022), and are also implicated in respiratory (Whiley & Hardie, 2015) and cardiovascular conditions (Chhibber-Goel et al., 2016; Teles & Wang, 2011), autoimmune dysfunction (Bruserud et al., 2018), diabetes (Casarin et al., 2013), obesity (Sen et al., 2017), Alzheimer's disease (Dominy et al., 2019; Miklossy, 2016), and cancer (Fan et al., 2018; Kato et al., 2016; Willis & Gabaldón, 2020). Since oral diseases are mediated by oral microbial ecology, or the oral microbiome, which

in turn is affected by human activity, understanding what drives the oral microbial community towards a state of pathogenicity is of great clinical significance.

The etiology of oral disease is multifactorial and not attributed to any singular pathogen but there is evidence of a correlation between host environment and disease (Marsh, 1994), firmly situating oral health within a one health framework. 'One health' is the notion that human health is interconnected with environmental and animal health (Huang et al., 2024). Human dietary practice not only dictates the nutrients available in an oral environment, but also impacts the lived environments with respect to land management, livestock proximity, and resource intensification (Silbergeld, 2019). Under certain environmental pressures, specific microbial taxa may stimulate oral dysbiosis i.e., the disrupted functionality of healthy microbial flora (Hajishengallis et al., 2012). Investigations of salivary and plaque biofilm communities have indicated that oral disease is associated with increased abundance of certain bacteria, such as *Tannerella forsythia, Treponema denticola*, and *Porphyromonas gingivalis* in the case of periodontal disease (Socransky & Haffajee, 2005), and *Streptococcus mutans* and others in dental caries (Kianoush et al., 2014; Loesche, 1986). Oral pathologies can be present in a paucity of these taxa, and conversely, these taxa may be present without progressive oral disease, suggesting that there is a more network-oriented effect in the development of pathogenicity (Lassalle et al., 2018; Nasidze et al., 2011). It is hypothesized that the transition of human societies towards intensive agricultural practices and industrialization influenced the progression of oral microbial communities towards a state of dysbiosis, resulting in increased incidences of oral disease across populations (Adler et al., 2013;

Cohen & Armelagos, 1984; Cordain et al., 2005; Dahlquist-Axe et al., 2024; Gancz et al., 2023).

Investigations of oral microbiomes throughout history is possible through the biomolecular study of dental calculus, a calcified plaque biofilm. A dense, crystalline matrix, calculus accumulates over an individual's lifetime and harbors more than 70% of its weight in microbial material (Adler et al., 2013; Warinner et al., 2014). Its accumulation is greater in populations with poor dental hygiene and high carbohydrate consumption, including those in antiquity that likely lacked oral care (Bos et al., 2019; Lieverse, 1999; Warinner et al., 2015; White, 1997). Biomolecules encapsulated within the calculus matrix are protected, to some extent, from environmental perturbation and taphonomic disruption, and thus, have been recovered from ancient human individuals across a wide range of geographic locations as well as from Neanderthals (Fellows Yates (A) et al., 2021). This temporal persistence and direct association with human remains renders dental calculus an ideal medium to investigate the evolution of human oral microbiomes, especially in the context of dietary transitions such as during the Neolithic.

Though ancient and historic populations often lack a written record of food consumed, possible diet can be elucidated from archaeological or isotopic evidence (Kennett et al., 2020; Prufer et al., 2021; Ray et al., 2024). However, inferring supposed diet from oral microbial composition requires an understanding of nuanced taxonomic and functional dynamics over large time scales. For instance, the divergence of modern humans from Neanderthals in the Middle Pleistocene Era was accompanied by a significant increase in host copy number of alpha-amylase genes, which function to break

down starches, and a concomitant increase in the relative abundance of *Streptococcus* species (Fellows Yates (B) et al., 2021). This likely stems from the increased reliance of anatomically modern humans on foods heavy in starch and the unique ability of *Streptococcus* species to bind host amylase. The relative consistency of the ecological niche of the human oral cavity following the divergence of anatomically modern humans from our closest relatives has resulted in relative stability in oral microbial biofilm communities across geographic and temporal scales (Fellows Yates (B) et al., 2021). Calculus accumulates on the surface of teeth over the life course of an individual barring any mechanical removal. Thus, broad compositional changes in oral microbiome composition have not been observed across Neolithic transitions (Fellows Yates (A) et al., 2021; Ottoni et al., 2021; Quagliariello et al., 2022). The adoption of agricultural practices by populations was an extensive process that took place gradually over many generations, likely contributing to the lack of significant compositional shifts. However, differential abundances of some taxa and the enrichment of some functional groups have been observed accompanying the adoption of agricultural practices (Modi et al., 2023; Ottoni et al., 2021; Quagliariello et al., 2022) and the industrialization of food production (Gancz et al., 2023; Ottoni et al., 2021). Network analyses have also revealed a shift in microbial communities, particularly after the incorporation of refined sugars and processed foods in human diets (Gancz et al., 2023).

Here, we present oral metagenomes recovered from the dental calculus of Ancestral Maya individuals from the unprecedented context of a single geographic area in the Americas occupied for over 10,000 years. Based on previously reported isotopic

analyses (Kennett et al., 2020) these individuals were hunters and gatherers with minimal cultivation until ~5,000 calibrated years before present (cal. BP) when they became committed maize farmers. This shift may have been influenced by migrating farmers from between 7,300 and 5,600 cal. BP (Kennett et al., 2022; Ray et al., 2024), with populations in this region transitioned to a diet comprising maize as a staple food source by 4,000 cal. BP (Kennett et al., 2020). Through shotgun metagenomic sequencing of calculus samples, we investigated changes in microbial community structure and diversity as well as evolutionary relationships of individual taxa across the adoption of maize agriculture in this region. Additionally, we evaluated how the functional potential of the oral microbiome corresponded to alterations in nutritional availability. This study offers an unprecedented view of oral microbial dynamics during the transition to agriculture preceding the formation of complex society in Mesoamerica.

2.3 Materials and Methods

2.3.1 Community engagement and ethics

All deceased humans in this study were excavated under permits issued by the Belize Institute of Archaeology (IA) and the Belize Forest Department. Skeletons of ancient individuals were exported under permits issued by the IA to K.M.P. with permissions to conduct molecular analyses on calculus and bone collagen. This research was conducted in close collaboration with the Ya'axché Conservation Trust, an internationally recognized Belizean NGO that is the co-manager of the Bladen Nature Reserve (BNR) with the

Government of Belize. Ya'axché is locally managed and largely staffed by members of descendent Maya communities. Over the course of this project, K.M.P. gave several consultation presentations to the local staff of Ya'axché and other interested community members on our research. In 2020, 2022, and 2024 in coordination with Ya'axché, we invited indigenous leaders and community members from villages proximate to the BNR to consult on this research. We provided advance invitations and arranged transportation up to 50 people from five villages to attend the public consultation. K.M.P., D.J.K., and S.J. have presented results of field and laboratory studies to the communities, and additional time was devoted to answering questions and clarifying the data and interpretations. Local participants were asked to provide feedback and ideas for future collaborative efforts and to be presented at future consultations. Community members requested future public consultations to update them on additional research results, as well as copies of all study results in English with translations into Mopan and Q'eqchi' languages.

2.3.2 Study population

The Ancestral Maya individuals included in this study (n=31) were excavated from two rock shelters located in the Maya Mountains of Belize (Kennett et al., 2020). These rock shelters, Mayahak Cab Pek (MHCP) and Saki Tzul (ST), are located within the Bladen Nature Reserve, a relatively isolated mountain valley (Prufer et al., 2021). Both rock shelters display stratigraphic sequences spanning ~12,000 to 1,200 cal. B.P., with consistent deposition of human bodies during that time. These sites are unique in that they are the only two in the lowland neotropics known to have been utilized across the

transition to agriculture. Individuals included in this study were chronologically organized based on direct radiocarbon dates and stratigraphic placement (Kennett et al., 2020; Ray et al., 2024). Stable carbon and nitrogen isotopic data from bone ($\boldsymbol{\delta}^{13}C_{col}$, $\boldsymbol{\delta}^{15}N_{col}$, and $\boldsymbol{\delta}^{13}C_{ap}$) and dentine ($\boldsymbol{\delta}^{13}C_{dentine}$ and $\boldsymbol{\delta}^{15}N_{dentine}$), were used to characterize the reliance on maize among these Ancestors. Although paleoecological data of this region suggest that maize was introduced as a food source around 6,500-5,500 cal. B.P., stable carbon and nitrogen isotopic analysis for individuals from these sites has indicated that those dating from 9,600 to 5,000 cal. B.P. (n=13) had a diet of primarily C_3 plants that characterize the neotropical forest, suggesting a "pre-maize" (PM) diet (Kennett et al., 2020; Ray et al., 2024). For individuals living between 5,000 and 4,000 cal. B.P. (n=3), isotopic values overlap with both pre- and staple-maize diets, suggesting a "transitional maize" (TM) diet. Isotopic data for individuals living after 4,000 cal. B.P (n=15) indicate that about 53% of total dietary carbon and greater than 50% of their dietary protein came from a C_4 plant or animals eating C_4 plants, suggesting a "staple maize" (SM) diet (Kennett et al., 2020).

2.3.3 Shotgun Sequencing of Dental Calculus

Dental calculus was sampled at the University of New Mexico. In cases where calculus was identified on multiple teeth from a single individual, sampled calculus was combined. Samples were sent to the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR) for shotgun sequencing and data analysis.

At LMAMR, calculus was subsampled and extracted following industry standards for ancient DNA (aDNA) (Dabney (A) et al., 2013; Orlando et al., 2021). When possible,

calculus was subsampled up to 10 mg to reduce destruction of original material. Calculus was irradiated in sterile Eppendorf tubes with the UV crosslinker (VWR) and washed with 0.5 M EDTA (pH=8.0) to remove surface contaminants. Samples were then incubated with 0.5 M EDTA (pH=8.0) and proteinase K (Qiagen) for at least 4 days. DNA was extracted using a modified silica column-based extraction protocol optimized for aDNA (Ozga et al., 2016) and eluted in EB buffer (Qiagen). An extraction blank was included in each round of DNA extractions to monitor for contamination.

DNA extracts were partially treated with the USER enzyme (uracil-DNA-glycosylase (UDG) and endonuclease VIII) followed by UGI, a UDG inhibitor, treatment (Rohland et al., 2015). This process excises most deaminated cytosines and cuts abasic sites but allows terminal nucleotides to retain damage typical of aDNA for authentication purposes (Briggs et al., 2010). Libraries were built using the blunt-end single-tube protocol established by Carøe et al. (Carøe et al., 2017) which optimizes the recovery of short fragments (< 70 bp). Double-stranded DNA was end-repaired, followed by blunt-end ligation of Illumina adapters. Each step was followed by heat inactivation rather than inter-reaction purification to reduce DNA loss (Carøe et al., 2017). Each batch of library preparation was accompanied by a library blank i.e. PCR-grade water (GeneMate) to monitor for contamination. Quantitative PCR (qPCR) was performed to confirm successful completion of library preparation, and the cycle threshold (Ct) value was used to infer the optimal number of cycles for indexing PCR. DNA libraries were dual-indexed in quadruplet with unique barcodes using the Kapa HiFi Uracil+ amplification enzyme (Kapa Biosystems). Libraries were cleaned with Sera-Mag SpeedBeads at a 1:1.8 ratio (Honap et al., 2023;

Rohland & Reich, 2012) and quantified with TapeStation (Agilent). Libraries were pooled based on DNA concentration and size selected to 150-500 bp using PippenPrep (Sage Systems). The libraries included in this study were shotgun sequenced across multiple paired-end 2 x 150 bp runs on the Illumina NextSeq and NovaSeq at the Oklahoma Medical Research Foundation (OMRF), Oklahoma City, US.

2.3.4 Data preprocessing

Some of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU). FASTQ files were processed using AdapterRemoval v2.3.3 (Schubert et al., 2016). These shotgun metagenomes had average reads of length 55-115 bp thus allowing for read merging. Reads were also trimmed of sequencing adapters and ambiguous bases (Ns) and reads with a Phred score below 30 and a length below 30 nucleotides were filtered. Collapsed and collapsed-truncated reads were concatenated and PCR duplicates were removed using fastp v0.23.4 (Chen, 2023) with adapter trimming and quality and length filtering disabled.

2.3.5 Human DNA assessment

Trimmed, merged, and deduplicated reads were assessed for the presence of human DNA through alignment to the complete human genome reference (T2T-CHM13v2). Reads were aligned using Bowtie v2.3.5.1 (Langmead & Salzberg, 2012) with default

parameters, and the resulting mapped and unmapped SAM files were processed with SAMtools v1.9 (Danecek et al., 2021). Unmapped reads were converted to FASTQ files with bam2FastQ from BamUtil v1.0.15 (Jun et al., 2015) and these filtered, “analysis-ready” or “AR” reads were used for further analysis. Duplicates were removed from the mapped bam files using DeDup v0.12.8 (Peltzer et al., 2016) and the resulting aligned, deduplicated BAM files were used as input for MapDamage v2.2.1 (Jónsson et al., 2013) to assess the rate of C-to-T and G-to-A transitions at the 5` and 3` ends. The BAM files were also filtered using PMDtools (Skoglund et al., 2014) to only include those reads with a postmortem damage score greater than 0, indicating a likely ancient origin, and then re-assessed with MapDamage2.

2.3.6 Taxonomic profiling

AR reads were classified through a dual approach to both assess prokaryotic taxonomic composition with MetaPhlAn4 and screen for eukaryotic DNA relevant for dietary reconstruction with Kraken2. MetaPhlAn v4.0.2 (Blanco-Míguez et al., 2023) uses a database of unique marker genes (mpa_vJan21_CHOCOPhlAnSGB_202103) to classify shotgun metagenomes to the species level (Blanco-Míguez et al., 2023). This method of classification is algorithmically similar to the functional annotation pipeline used in this analysis, HUMAnN v3.6 (Beghini et al., 2021), and thus was retained for compositional data analysis. This MetaPhlAn version includes species-level genome bins created from metagenomically assembled genomes from various body sites, geographic locations, ages, and lifestyles, increasing the presence of understudied or uncharacterized species in

the database (Pasolli et al., 2019). MetaPhlAn was run with a minimum read length of 30 and the unclassified estimation was retained.

Metagenomic samples were also classified using Kraken v2.1.2 (Wood et al., 2019) against a custom database of representative bacteria and archaea from the Genome Taxonomy Database (GTDB) and representative fungi, protozoa, and viruses, as well as mitochondria and plastids, from the National Center for Biotechnology Information reference sequence database (NCBI RefSeq). Kraken relies on exact *k*-mer matching to a library of genomes to obtain the lowest common ancestor (LCA). The accompanying tool, Bracken v2.0 (Lu et al., 2017), was then used to re-estimates the relative abundances to a specified taxonomic level based on the composition of the dataset (Lu et al., 2017). The ability to customize the search database allowed us to screen for direct evidence of dietary genetic material through a competitive mapping strategy. The Kraken database for the present analysis was built, and samples were classified with default parameters. The bracken database was built and profiles re-estimated with a *k*-mer size and read length of 35.

2.3.7 Ancient authentication and source tracking

Ancient DNA has characteristically short read lengths (< 300 bp) due to hydrolytic depurination and beta elimination, resulting in lengths typically falling within the 50-150 bp range (Dabney (B) et al., 2013; Orlando et al., 2021). Additionally, the terminal nucleotides of aDNA fragments tend to accumulate the hydrolytic deamination of cytosine residues, resulting in an apparent cytosine to thymine (C to T) or guanine to adenine (G to A)

misincorporation at the 5` and 3` ends, respectively (Dabney (B) et al., 2013). These patterns of degradation are used to authenticate the nature of DNA reads in metagenomes. Here, mapDamage plots from alignment to the most abundant oral microbes (Table 2.1) present in the samples were assessed for C-to-T or G-to-A transitions at the terminal ends.

Species	Strain	NCBI TaxID
Anaerolineaceae bacterium oral taxon 439	W11661	1889813
Methanobrevibacter oralis strain	YH	66851
Eubacterium minutum	ATCC 700079	888721
Actinomyces dentalis	DSM 19115	272548
Unclassified Firmicutes (GGB51111_SGB71327)	VelskolM_2019__ERR3003648__bin.5	NA

Table 2.1 Most abundant oral taxa.

SourceTracker v2.0.1 (Knights et al., 2011) was used to assess the likely source environments of reconstructed metagenomes. Briefly, sourcetracker2 estimates the proportion of contributing environments through Bayesian estimation using taxonomic profiles. Source tracking was performed on profiles classified by both MetaPhlAn4 (Blanco-Míguez et al., 2023) and Kraken2 (Wood et al., 2019). The sample profiles produced by the classifiers were the sinks that were assigned predicted environmental proportions based on taxonomic profiles of source samples (Knights et al., 2011). These sources included previously published shotgun metagenomic data from modern dental calculus (Velsko et al., 2019), sub- and supra-gingival plaque (HMP, 2012), feces from industrialized and non-industrialized societies (Obregon-Tito et al., 2015; Rampelli et al., 2015), soil (Johnston et al., 2016), and skin (Oh et al., 2014). All source and sink profiles from each tool were merged, relative abundance was re-calculated to output absolute

abundances based on number of reads processed, and merged abundance tables were summarized to either the genus or species level. Sourceracker2 gibbs was run on species and genus tables, rarefying source and sink profiles to 30,000 reads per sample and including a prior species count in training environments (alpha1) of 0.01 relative to the number of sequences and a prior count in unknown environments (alpha2) of 1.0 relative to the input. These two values prevent unlikely taxa in source environments from influencing the classification and prevents overfitting of the unknown environment (Gancz et al., 2023; Knights et al., 2011).

2.3.8 Functional profiling

Ancestral Maya metagenomes were characterized for functional capacity using HUMAnN v3.6 (Beghini et al., 2021). HUMAnN3 taxonomically classifies metagenomes and, with a database of identified pangenomes, annotates the functional capacity and source taxa of the input sequences (Beghini et al., 2021). HUMAnN3 was run against the ChocoPhlAn v201901_v31 database with the uniref50 search mode and a minimum alignment length of 30 for MetaPhlAn classification. The resulting gene family tables were joined and re-normalized from their reported reads per kilobase (RPK) to copies per million (CPM) to account for differences in sequencing depth. Re-normalized tables were then regrouped into KEGG Orthogroups (KOs) (Kanehisa & Goto, 2000) to categorize gene families into higher order metabolic functional groups and observe how these were affected by the dietary transition (Kanehisa et al., 2014).

2.3.9 Statistical comparisons

Statistical analyses for all metrics were performed in R v4.2.2. MetaPhlAn output was restricted to oral taxa, including those in the human oral microbiome database (HOMD) (Escapa et al., 2018), those with an isolation source indicated on GTDB or NCBI as oral or respiratory, or those identified through source tracking to belong to modern human calculus or plaque (HMP, 2012; Velsko et al., 2019). Oral profiles were filtered of rare taxa, retaining those preset at greater than 0.005% relative abundance in at least 20% of samples. The filtered table, sample metadata and the phylogenetic tree appropriate for the MetaPhlAn marker database version (mpa_vJan21_CHOCOPhlAnSGB_202103.nwk) were loaded into phyloseq v1.42.0 (McMurdie & Holmes, 2013). Microbiome data are compositional, where individual abundances are relational to the whole and subject to discrepancies when filtering rare taxa. To accommodate this, a pseudocount of 0.001 was added to all taxonomic profiles to account for zeroes and data were center log-ratio (CLR) transformed (Gloor et al., 2017). A principal component analysis (PCA) was performed to reduce dimensionality of the data and observe variation between metadata variables. Principal components were ordinated using the mixOmics v6.22 (Rohart et al., 2017) package. Permutation analysis of variance (PERMANOVA) testing was performed on Euclidean distances of PC components to explore the statistical significance of subsistence pattern, archaeological site, number of AR reads, and extraction and library batches using adonis2 from the vegan v2.6-4 package (Oksanen et al., 2001). A PCA bi-plot with species correlating to the top five positive and negative PCA loadings was visualized using ggplot2 v3.4.3 (Wickham, 2016). To measure homogeneity of group dispersions and

ensure any statistical significance observed was not due to differences in the spread of principal components, the betadisper command from the vegan package was performed on Euclidean distances, followed by an analysis of variance (ANOVA) using the compositions v2.0-8 library (http://www.stat.boogaart.de/compositions/).

Merged, normalized, and regrouped KO gene family tables stratified to include taxonomic contributions were loaded into R and filtered to include KOs contributed by oral taxa and present at > 2 CPM in at least 20% of samples. Stratified output was aggregated to explore taxonomic contribution to functional capacity. The resulting table was loaded into a phyloseq object along with sample metadata and the tree file appropriate for version mpa_v30_CHOCOPhlAn_201901, the most recent version for HUMAnN3. Functional profiles are also compositional in nature and thus were transformed as described above for taxonomic profiles but with a pseudocount of 1 added prior to CLR transformation. Subsequent processes were identical to that described for MetaPhlAn4 taxonomy tables to generate a PCA plot and perform PERMANOVA testing on Euclidean distances between principal components.

Taxonomic and functional profiles were queried for differential abundances across the transition to maize agriculture. Log fold change was measured for each species in the oral filtered MetaPhlAn4 table and a Pairwise Wilcoxon rank sum test with Benjamini-Hochberg correction for false discovery rate (Benjamini & Hochberg, 1995) was performed for each. Functional profiles were distributed into metabolically specific tables based on KEGG categories for KO groups. Metabolic categories were obtained through mapping files provided by HUMAnN legacy data (Abubucker et al., 2012). Functional categories related to

acid tolerance and resistance (ATR) mechanisms described in (Cotter & Hill, 2003; Foster, 2004; Guan & Liu, 2020; Iyer et al., 2002; Krulwich et al., 2011; Lund et al., 2014; Ma et al., 2003) were collected from both KO and UniRef50 functional annotations. Differential abundance for both taxa and functional annotations were measured with log-fold-change values and tested by Pairwise Wilcoxon rank sum tests with Benjamini-Hochberg correction as described above.

2.3.10 Genome reconstruction and phylogeny

Reference-based mapping and variant calling

AR reads were aligned to the NCBI reference genomes of oral taxa *Pseudoramibacter alactolyticus*, *Tannerella forsythia*, and *Streptococcus sinensus* using BWA aln v0.7.17 (Li & Durbin, 2009) with seed disabled and -n 0.1. Output SAM files were converted to BAM format, indexed using SAMtools, and deduplicated using DeDup with the –m parameter. MapDamage2 was used to measure the rate of terminal nucleotide misincorporation and to rescale the alignment file to account for damaged positions. The trimBam command from BamUtil was used to trim the two terminal nucleotides at the 5` and 3` positions. Variant calling was performed using VarScan v2.3.9 (Koboldt et al., 2009) with a minimum coverage of 5, a minimum of 3 reads and 20% of reads covering a variant allele to call it a SNP, a minimum average quality of 37, a minimum of 90% of reads covering a variant allele to call it a homozygous SNP, a p-value of 1, and a strand filter of 0. Alignments to *T. forsythia* and *P. alactolyticus* with a heterozygosity ratio < 1 and > 30% of the reference covered five times were retained for phylogenetic analysis. Only one

alignment to *S. sinensis* had a heterozygosity ratio < 1, but it only covered 3% of the reference at a depth of five times and was thus excluded from phylogenetic reconstruction.

Tannerella forsythia whole-genome enrichment

Libraries from seven samples, and their associated blanks, were selected for whole-genome capture of *Tannerella forsythia*. DNA was isolated from *T. forsythia* ATCC43037 using the DNEasy Blood and Tissue Kit (Qiagen) following the manufacturer's protocol. 5 µg of DNA was used to prepare RNA baits covering the entire genome at Daicel Arbor Biosciences. The enrichment was carried out following the manufacturer's protocol (MyBaits v4) and a hybridization temperature of 65°C was used. Two rounds of enrichment were conducted. After each round, enriched libraries were amplified following manufacturer's protocol and using the KAPA HiFi HotStart kit (KAPA Biosystems), with the 30 µL of captured library being amplified over two replicates for 14 cycles. Amplified enriched libraries were purified using the SpeedBead cleanup (1.8X), assessed using the TapeStation D1000 assay, and pooled in equimolar amounts. The resulting sequencing pool was sequenced on the Illumina NovaSeq SP paired-end 2 x 150 bp run at OMRF. Shotgun reads were trimmed and merged using AdapterRemoval v2.3.3 (Schubert et al., 2016), AR reads were aligned to the *T. Forsythia* reference strain 92A2 and variant sites were called as described above.

Phylogenetic reconstruction

For phylogenetic reconstruction, comparative ancient and modern dental calculus metagenomes from the Ancient Metagenomic Directory (https://github.com/SPAAM-community/AncientMetagenomeDir) (Fellows Yates (A) et al., 2021) were screened for the genomes of interest in the same manner described above. Strains and metagenomically assembled genomes (MAGs) published on NCBI RefSeq were also included in phylogenetic reconstruction. A list of comparative reconstructed strains is provided in Supplementary Table 2 and Supplementary Table 3. Reference genomes were simulated into paired-end reads using wgsim (Li, 2011) with parameters described in (Honap et al., 2023). Reads were merged, aligned to their respective reference, and variants were called in the same manner as the other samples. All VCF files were merged, indexed, and filtered to biallelic SNPs using BCFtools v1.17 (Danecek et al., 2021). VCFtools v0.1.17 (Danecek et al., 2011) was used to exclude genomic regions containing repetitive, mobile, RNA-, or phage-related elements described previously for *P. alactolyticus* (Fleskes et al., 2024) and *T. forsythia* (Honap et al., 2023). Insertion-deletion and SNP positions were removed if they had < 80% coverage for *P. alactolyticus* and < 90% for *T. forsythia*. The final VCF file was converted to a multi-fasta file and used as input for IQTree with the following parameters: -m MFP+ASC -bb 1000 -alrt 1000 -bnni. Trees were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.4. Results

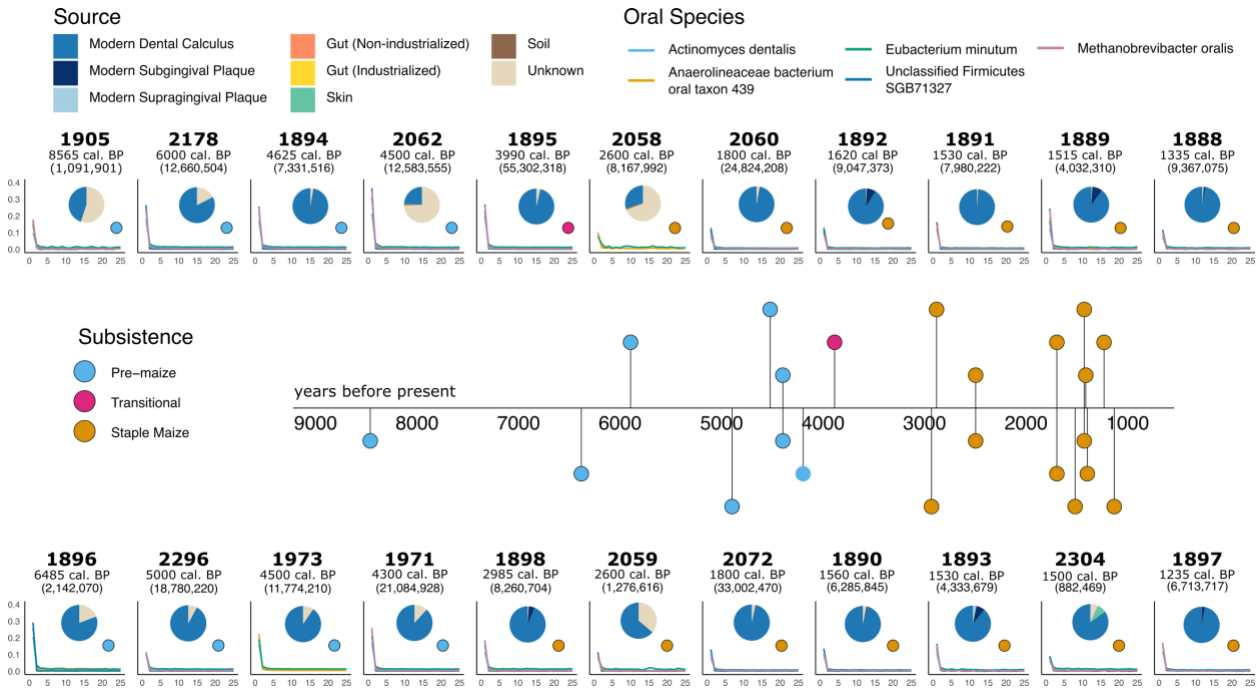


Figure 2.1 Visual summary of the Ancestral Maya metagenomes.

Pie charts indicate proportion of metagenome attributed to an oral source. They are situated within a nucleotide misincorporation plot indicating C to T transitions at the terminal 5` nucleotides for 5 most abundant oral microbes. Timeline is approximated based on calibrated ^{14}C radiocarbon dates. Subsistence pattern is indicated by the colored circle in the lower right-hand side of each sample plot. Number of AR reads for each sample indicated under radiocarbon date in parenthesis.

2.4.1 Shotgun Sequence Processing and Authentication

Of the 31 individuals from which dental calculus was sampled, oral metagenomes were successfully recovered from 22 individuals from two archaeological sites spanning over 7,000 years. Double-stranded, partially uracil-DNA-glycosylase-treated libraries were shotgun-sequenced and resulted in 2,823,718 to 75,282,291 raw reads per library. Details can be found in Supplementary Table 4. Paired-end sequences were trimmed and merged, and PCR duplicates were removed, resulting in 1,591,879 to 67,466,335 quality-filtered, merged reads. The duplication rate for the samples ranged from 0.94% to 25.45%.

Trimmed, merged, and deduplicated reads were aligned to the human reference genome (the telomere-to-telomere complete genome, T2T-CHM13v2.0) to determine endogenous host DNA content. After filtering the alignment to include only those with postmortem damage (PMD), all samples had a rate > 1% of C to T or G to A nucleotide transitions at the 5` and 3` ends (Green et al., 2010; Rohland et al., 2015), suggesting the presence of ancient human DNA in the dental calculus samples. Metagenomes from two individuals (MRSC-2304 and MRSC-2076) had more than 10,000 reads aligning to the reference genome with a mean coverage of 0.0264X and 0.0111X, respectively. Alignment to the human mitochondrial reference genome (Revised Cambridge Reference Sequence, rCRS) and subsequent PMD filtering resulted in insufficient genome coverage to determine mitochondrial haplogroups. A table outlining summary statistics for alignments to both T2T-CHM13v2.0 and rCRS are given in Supplementary Table 5. Reads that aligned to the human genome were removed for downstream analyses, and the remaining filtered reads are hereafter referred to as “analysis-ready (AR) reads”.

Microbial source tracking was employed to ascertain the proportion of metagenomic reads attributed to an oral source compared to other possible contaminating sources. A comparative analysis of classification algorithms, Kraken2 (Wood et al., 2019) and MetaPhlAn4 (Blanco-Míguez et al., 2023) was performed for source tracking. For the known sources, both programs classified reads from oral metagenomes at relatively equal rates, but Kraken2 classified a greater proportion of reads from skin, soil, and fecal environments. As a result, Kraken2 output was used for source tracking. A comparison of the classification rates can be seen in Figure 2.2. Of the 31 metagenomes

assessed, one had too few reads (< 30,000) classified by Kraken to be included in the source tracking analysis and six did not display an oral signature (< 25% reads classified as oral).

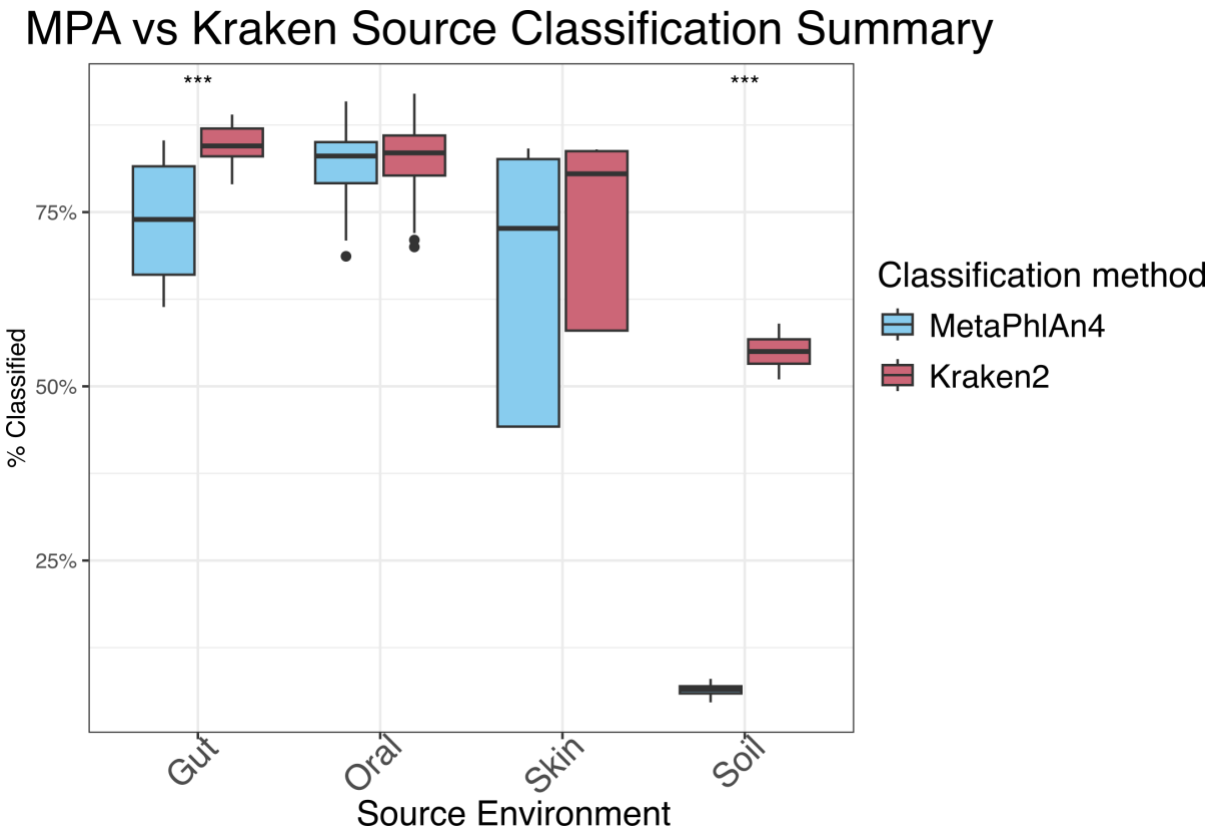


Figure 2.2 Rate of classification for different sources between MetaPhlAn4 and Kraken2.

Modern calculus, supra-, and subgingival plaque were combined for oral environments and feces from industrial and non-industrial populations were combined for the gut environment. Plot generated in R using ggplot2 (Wickham, 2016). Asterisks indicate statistical significance (*** $p < 0.001$) with Pairwise Wilcoxon test.

Metagenomes that displayed an oral signature were aligned to the most abundant oral microorganisms to determine ancient origin by measuring the frequency of C-to-T and G-to-A transitions at the 5` and 3` ends, respectively. This includes *Actinomyces dentalis*, Anaerolineaceae bacterium oral taxon 439, *Eubacterium minutum*, an unclassified

Firmicutes species (SGB71327), and *Methanobrevibacter oralis*. The frequency of C-to-T transitions ranged from 8.1-31.7% and that of G-to-A transitions ranged from 7.8- 32.4% for all metagenomes. This pattern of degradation is congruent with aDNA (Green et al., 2010; Orlando et al., 2021). Of the 24 metagenomes with ancient oral signatures, two had fewer than 500,000 AR reads, and because such low sequencing depth could preclude capturing available species diversity, these metagenomes were removed from further analyses. Ancient oral metagenomes were successfully recovered from 8 out of the 13 PM individuals, 1 out of the 3 TM individuals, and 13 out of the 15 SM individuals. Raw sequence data from these 22 metagenomes filtered of reads mapping to the human genome, as well as paired-end raw *Tannerella forsythia* sequence capture data, were deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1257829. These groups of divergent subsistence strategies allowed us to investigate differential microbial abundances and compositions in response to host lifestyle transitions.

2.4.2 Oral Microbial Taxonomic Composition

Reconstructed metagenomes were taxonomically classified using two distinct methodologies: 1) through MetaPhlAn4 (Blanco-Míguez et al., 2023), reads were aligned to their curated clade-specific marker database, and 2) through Kraken2 (Wood et al., 2019), segments of reads of length *k* were matched to *k*-mers of a custom library of prokaryotic and eukaryotic genomes, plastids, and mitochondria. MetaPhlAn4 is our primary means of classification for bacterial and archaeal taxa in oral microbiomes. The algorithm automatically computes relative abundance normalized to accommodate genome lengths

and incorporates metagenomically assembled genomes (MAGs) to expand available microbial diversity (Blanco-Míguez et al., 2023). Furthermore, the use of species-specific marker genes results in a low false-positive rate relative to other classification tools regularly employed by aDNA researchers (Velsko et al., 2018). Classification with Kraken2 requires a normalization step and can be computationally intensive when using an inclusive reference database. However, the customized database used in this study included mitochondrial and chloroplast genomes to screen for DNA that would be informative of dietary constituents but excluded whole eukaryotic genomes to decrease computational requirements. The inclusion of the GTDB representative genome database encouraged a competitive mapping strategy to reduce false positives. This dual approach allowed for the discovery of potential eukaryotic DNA that would provide evidentiary value of dietary constituents through Kraken2, while also producing a taxonomic profile with MetaPhlAn4 through the same marker-gene based approach utilized by the functional profiling tool, HUMAnN3 (Beghini et al., 2021).

MetaPhlAn 4

The percent of metagenomic reads classified at the species level using MetaPhlAn4 ranged from 10.84% to 76.76%. The most abundant organisms identified included Anaerolineaceae bacterium oral taxon 439, *Methanobrevibacter oralis*, and an unclassified Firmicutes taxa (SGB71327) metagenomically assembled from modern dental calculus by Velsko and colleagues (2019). Of the 307 species identified, 155 were present at greater than 0.005% in at least 20% of samples and were confirmed to be oral microbes. Due to

the compositional nature of microbiome data, the filtered table was transformed by its center log-ratio (CLR) (Gloor et al., 2017; Quinn et al., 2019). A heatmap that includes the 30 most abundant species in each subsistence group is depicted in Figure 2.3.

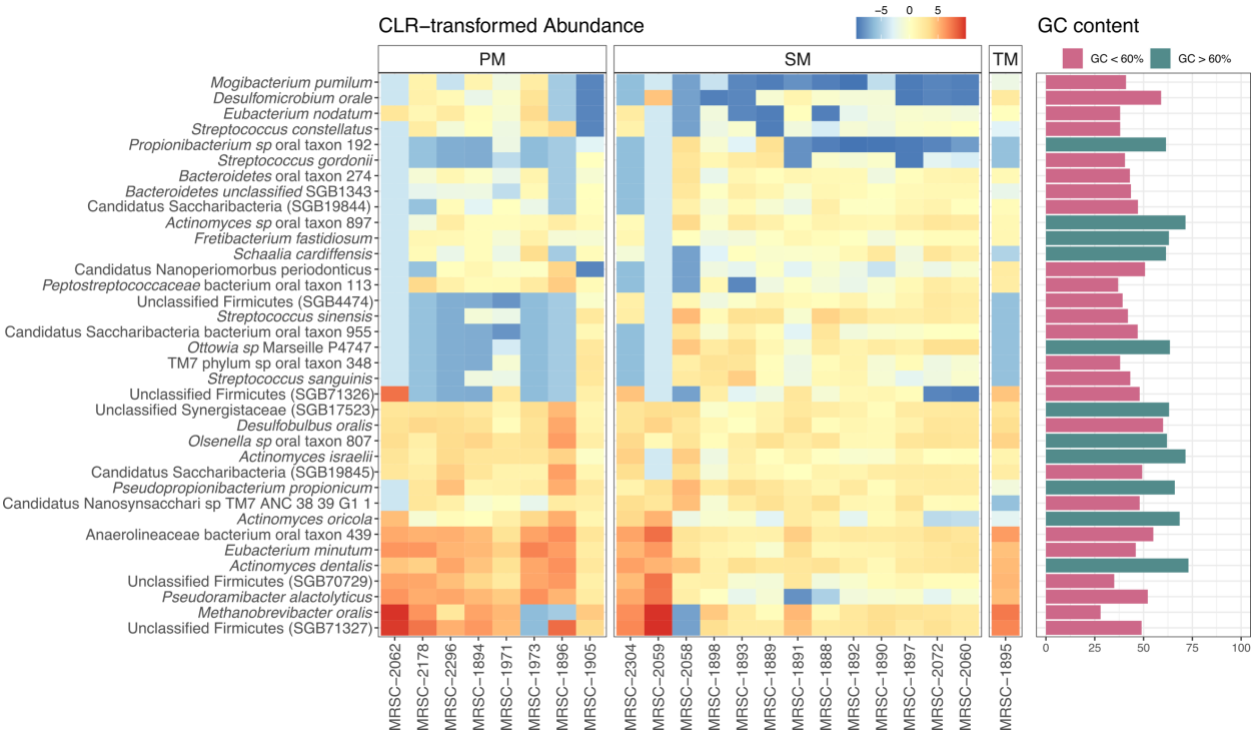


Figure 2.3 Most abundant taxa with accompanying GC content.
Heatmap depicts CLR-transformed abundance values of most abundant taxa present across subsistence strategies. GC content for each taxon is depicted to the right.

To reduce the complexity of the filtered abundance table and observe variance among the data, a principal component analysis (PCA) was conducted on the CLR-transformed table (Figure 2.4) (Gloor et al., 2017; Quinn et al., 2019; Ramette, 2007). A permutated analysis of variance (PERMANOVA) revealed no significant influence by burial site, number of AR reads, Cq value, library batch, or extraction batch. With respect to subsistence pattern, the TM calculus sample clustered with PM samples in negative PC1 space (Figure 2.4A). However, this sample was excluded for statistical testing for

subsistence pattern distribution. After its exclusion, subsistence pattern (R^2 = 0.12, F = 2.57, p = 0.010) significantly contributed to principal components and had a non-significant heterogeneity of variance indicating that the influence of subsistence pattern on ordination was not due to differences in dispersal (Anderson, 2006). To visualize the taxa contributing to PC ordination, a biplot of PCA loadings was generated onto the PCA plot limited to SM and PM metagenomes (Figure 2.4B). SM metagenome dispersion, distributed across PC1 but almost exclusively represented in positive PC1 space, was driven by two *Streptococcus* species, candidate division TM7, *Neisseria sicca*, and *Ottowia* sp. Marseille P4747 (Figure 2.4B). PM metagenome dispersion clustered tightly in the top left quadrant and was largely driven by *Pseudoramibacter alactolyticus*, two *Eubacterium* species, *Treponema lecithinolyticum*, and an unclassified Firmicutes species (SGB70729).

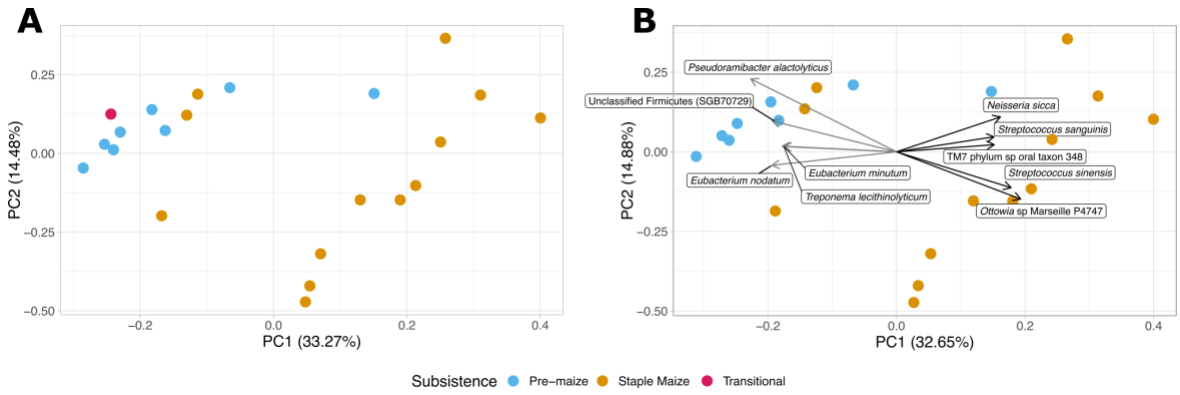


Figure 2.4 PCA plot of CLR-transformed relative abundances from MetaPhlAn4.
Metagenomes were filtered to > 0.005% relative abundance in at least 20% of samples and only oral taxa are retained. Samples are colored by subsistence pattern. A) PCA includes TM calculus sample. B) TM calculus sample is excluded and top 5 loadings affecting sample distribution are represented.

Some of these taxa driving dispersion were differentially abundant either before or after the adoption of a SM diet. The relative abundance of the Clostridiales species, *P.*

alactolyticus, *E. minutum*, and the unclassified Firmicutes (SGB70729), significantly decreased after the adoption of maize, while *N. sicca*, *S. sinensis*, *S. sanguinis*, and *Ottowia* sp. Marseille P4747 significantly increased. Species enriched after the adoption of maize belong to clades that require a mixed community for growth, including candidate Saccharibacteria and their obligate symbiont, *Actinomyces* species (Figueroa-Gonzalez et al., 2020; Naud et al., 2023). Differential taxa can be seen in Figure 2.5.

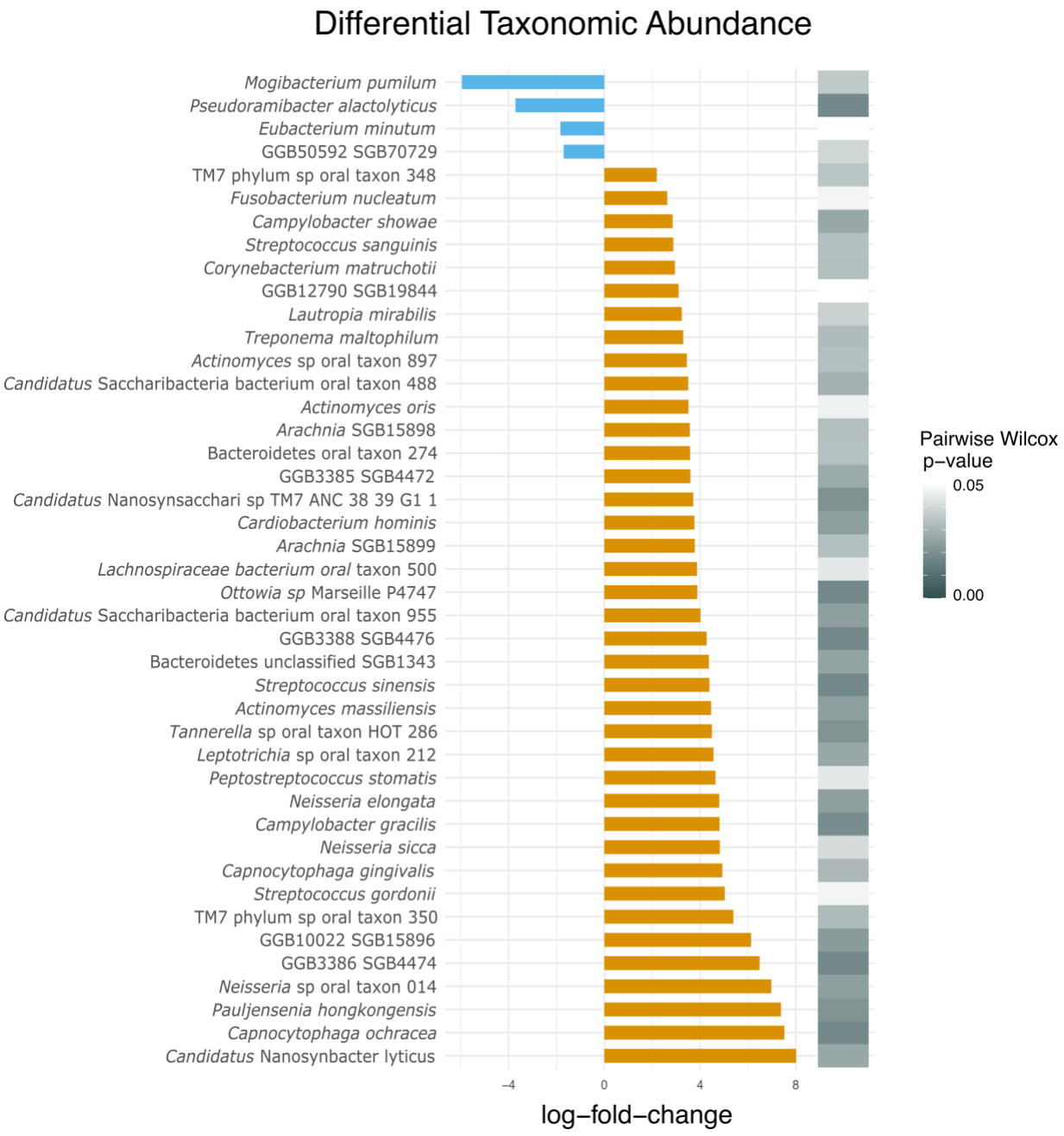


Figure 2.5 Log-fold-change of microbial species across agricultural shift and Pairwise Wilcoxon tests for significance.

Includes species with a Benjamini-Hochberg FDR-corrected p-value near or below a significance threshold of $p < 0.05$. Color gradient depicts p-value significance. p-values above 0.05 have a white square.

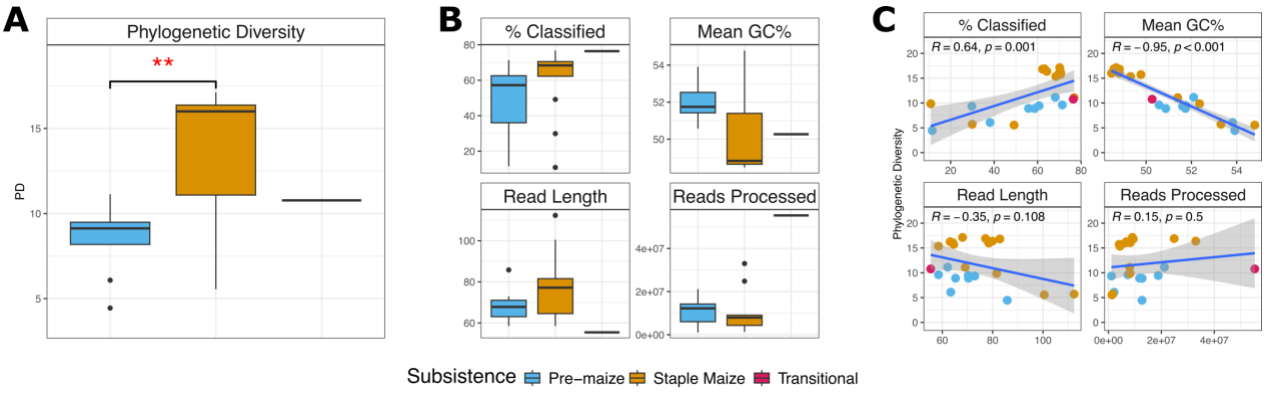


Figure 2.6 Alpha diversity and other metrics

Phylogenetic diversity (PD) across subsistence strategy shift compared to other measures, including mean GC content, number of reads processed, average DNA read length, and average classification rate across subsistence strategies. A) PD compared across subsistence strategies. Pairwise-Wilcoxon test (* p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001). B) Other metagenome metrics compared across subsistence strategies. No significant differences identified. C) Linear correlation between different metrics and PD. Pearson correlation (*R*) indicated on each plot.

There was a significant shift in intra-individual species diversity from PM to SM diet. Faith's Phylogenetic diversity (PD) (Kruskal-Wallis test, p = < 0.01) and species richness (Kruskal-Wallis test, p = < 0.01) were both significantly higher in individuals with a SM diet than a PM diet (Figure 2.6A). There were no significant differences with respect to mean GC content, the number of reads processed with MetaPhlAn4, the average DNA read length of reads processed, nor the rate of classification across subsistence strategies (Figure 2.6B). A linear regression of these metrics as they pertain to PD indicated that classification rate positively (R=0.64, p < 0.01) and mean GC content negatively (R=-0.95, p < 0.01) correlates to PD while read length and number of reads processed shows no correlation.

2.4.3 Dietary assessment

Given that maize was a dietary staple for the SM individuals, we directly screened for evidence of *Zea mays* genetic material in the metagenomes of these individuals. To shed light on microbial dynamics of a dietary shift given direct isotopic evidence for maize consumption, we characterized the functional capacity of the oral microbiomes. We focused on functions related to carbohydrate metabolism due to the knowledge that about half the diet of the SM individuals was attributed to maize. Furthermore, carbohydrate fermentation can reduce the pH of an environment, so we investigated markers for acid tolerance and resistance.

Eukaryotic DNA signature

To determine if dietary plant DNA was present in the Ancestral Maya dental calculus, metagenomes were classified with a custom Kraken database that includes eukaryotic mitochondrial and plastid genomes, as well as viral, fungal, archaeal, and bacterial genomes to encourage a competitive mapping strategy. This resulted in a total of 526 eukaryotic species identified in at least one metagenome prior to any filtering measures. There should be at least 500 reads classified as eukaryotic to attempt the validation process (Mann et al., 2023), and the majority of those species classified in these metagenomes fall below this threshold. The most abundant dietary plant DNA present in detectable amounts was found in dental calculus from two individuals (MRSC-1895, TM; MRSC-2062, PM) and belonged to *Triticum aestivum*, or common wheat. This was not present in Mesoamerica prior to European contact and thus is a false positive. *Z. mays*, or

maize, was not identified in any metagenomic samples. However, after alignment to the *Z. mays* genome using the Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009), 18 calculus samples had > 500 unique reads aligning to the reference and 16 had > 1000. *Z. mays* was detected in one metagenome (MRSC-1895, Transitional subsistence) at > 8000 reads and the frequency of C to T and G to A transitions suggest possible ancient origin.

	MRSC-1895
Analysis-ready reads	55302318
Unique mapped reads	8240
Average length of mapped reads	43.109
Mean coverage	0.0002X
Std dev of mean coverage	0.0654X
% reference covered >= 1X	0.01%
% endogenous DNA	0.0001

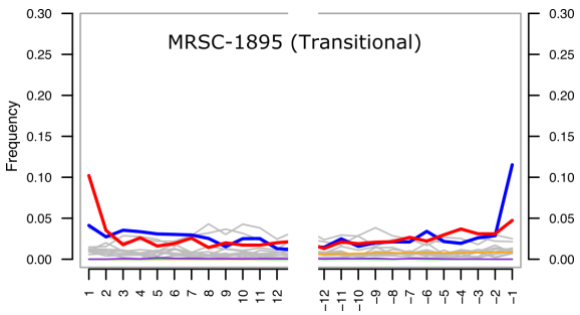


Figure 2.7 *Zea mays* alignment summary for MRSC-1895.

Functional composition

Metagenomes were characterized for functional potential with HUMAnN3 (Beghini et al., 2021) using the UniRef database clustered at 50% identity (UniRef50) (Suzek et al., 2015). The gene family output tables were combined and normalized to copies per million (CPM) to account for differences in sequencing depth. Gene families were then regrouped into KEGG orthogroups (KO) and categorized based on metabolic function. KEGG orthologs are defined from experimentally validated functions within specific species and generalized based on sequence similarity to provide a high-level functional description from genomic or protein sequences (Kanehisa et al., 2016). This reconfiguration from protein reference clusters with minimized redundancy (UniRef50) (Suzek et al., 2015) to a higher level functional annotation with KEGG allows for a more network-oriented

resolution to metagenomic function. The KO table was filtered to include only oral taxa with KOs present at greater than 2 CPM in at least 20% of samples. The filtered table was CLR transformed and ordinated in a PCA to observe compositional patterning and reveal any variation due to metadata. As with the taxonomic composition, a PCA of KO functional groups revealed a significant separation (PERMANOVA; $p < 0.05$) with respect to both subsistence pattern and archaeological site. However, when controlling for either variable, only subsistence pattern remained significant. The functional table was aggregated to species level to generate top species loadings for the PCA plot (Figure 2.8)

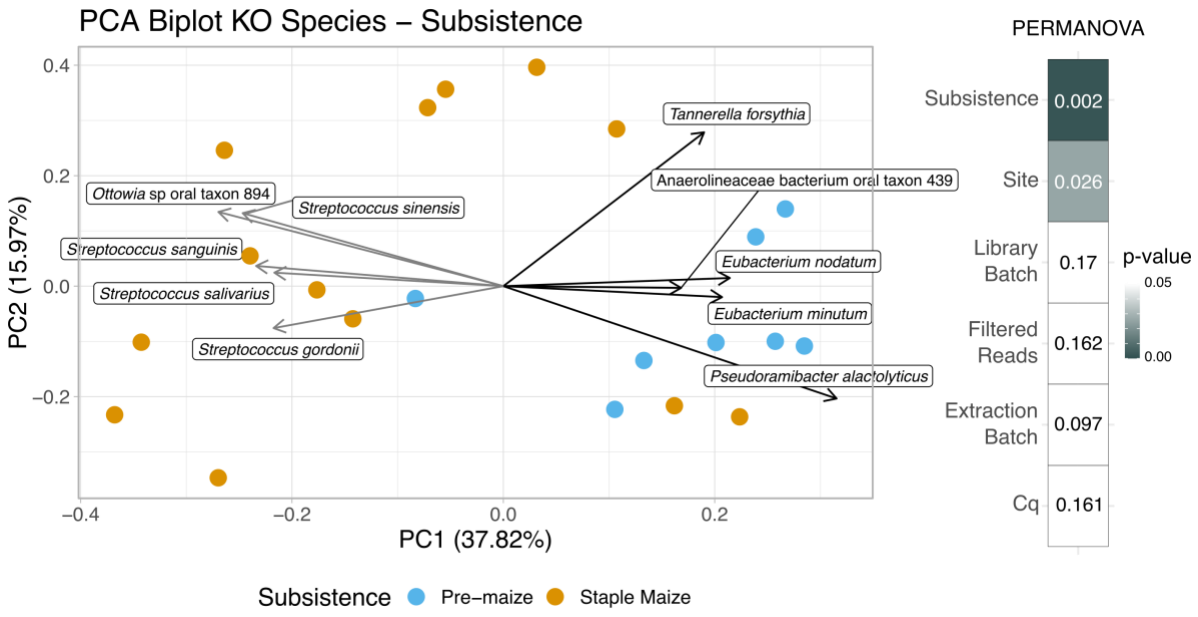


Figure 2.8 PCA bi-plot of taxa contributing to functional potential.

Gene families were normalized to CPM and regrouped into KEGG orthologous groups. Data was filtered to greater than 2 copies per million in at least 20% of samples and only includes oral taxa. Level of significance of each metadata category determined by PERMANOVA analysis.

At the time of this writing, the pangenome database available for HUMAnN v3.6 (Beghini et al., 2021) functional profiling excludes species level genome bins (SGBs)

present in the MetaPhlAn v4.0.2 database and thus taxonomic patterns observed here lack this nuance. With this caveat, the taxa indicated as top loadings contributing to separation across PC space differ slightly from those represented by MetaPhlAn4. In addition to *Streptococcus sinensis, S. sanguinis,* and *S. gordonii* driving the separation of SM functional profiles across PC space, *S. salivarius* also contributes significantly. Three different *Streptococcus* species (*S. sinensis*, *S. sanguinis*, and *S. salivarius*) contribute a significantly greater abundance of metabolic functions in the SM metagenomes. *E. minutum* and *P. alactolyticus* predominantly contribute to metabolic function in PM metagenomes and, along with Anaerolineaceae bacterium oral taxon 439 and *E. nodatum*, drive the separation of PM functional potential in PCA space.

In almost all metabolic categories, the relative abundance of functions increases on average in SM diets with the exception of biosynthesis of secondary metabolites (Figure 2.9). Significant shifts in abundance across subsistence strategies was determined by a Pairwise Wilcoxon rank sum test with Benjamini-Hochberg false discovery rate (FDR) correction (Benjamini & Hochberg, 1995). There was an increase in relative abundance of glycan biosynthesis and metabolism, carbohydrate metabolism, metabolism of cofactors and vitamins, and metabolism of terpenoids and polyketides, nucleotide metabolism, amino acid metabolism, energy metabolism, lipid metabolism, and KO values involved in multiple metabolic processes at a significance of adjusted $p < 0.001$. To observe any patterns of inter-group species diversity (beta diversity) between subsistence patterns with respect to metabolic category, a PERMANOVA analysis was performed on Euclidean distances of CLR-transformed relative abundances of taxa contributing to metabolic

function. Subsistence pattern significantly contributed to this diversity ($p < 0.05$) for all metabolic categories except energy metabolism, and a test of beta dispersion indicated that statistical significance was not due to heterogeneity of variances. Furthermore, phylogenetic diversity and species richness was significantly higher in SM metagenomes.

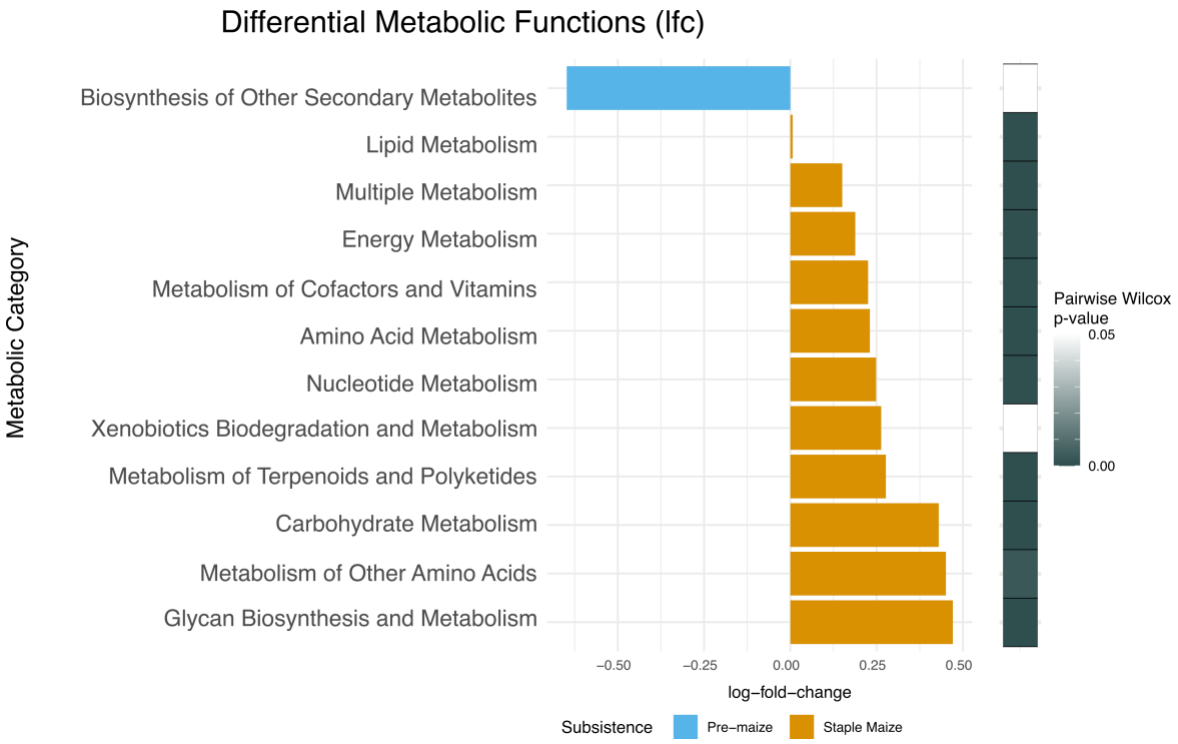


Figure 2.9 Metabolic shifts across dietary transition.

Log-fold-change of metabolic function determined by KO value. Pairwise Wilcoxon test performed with FDR correction.

To obtain a more detailed understanding of the distribution of functions among taxa, the differential abundances of taxa contributing to each category was quantified. Figure 2.10 details each metabolic category and the taxa that significantly change across subsistence strategies. *P. alactolyticus* and *E. minutum* are consistently enriched in PM metagenomes, while the taxa enriched in SM metagenomes are numerous and diverse.

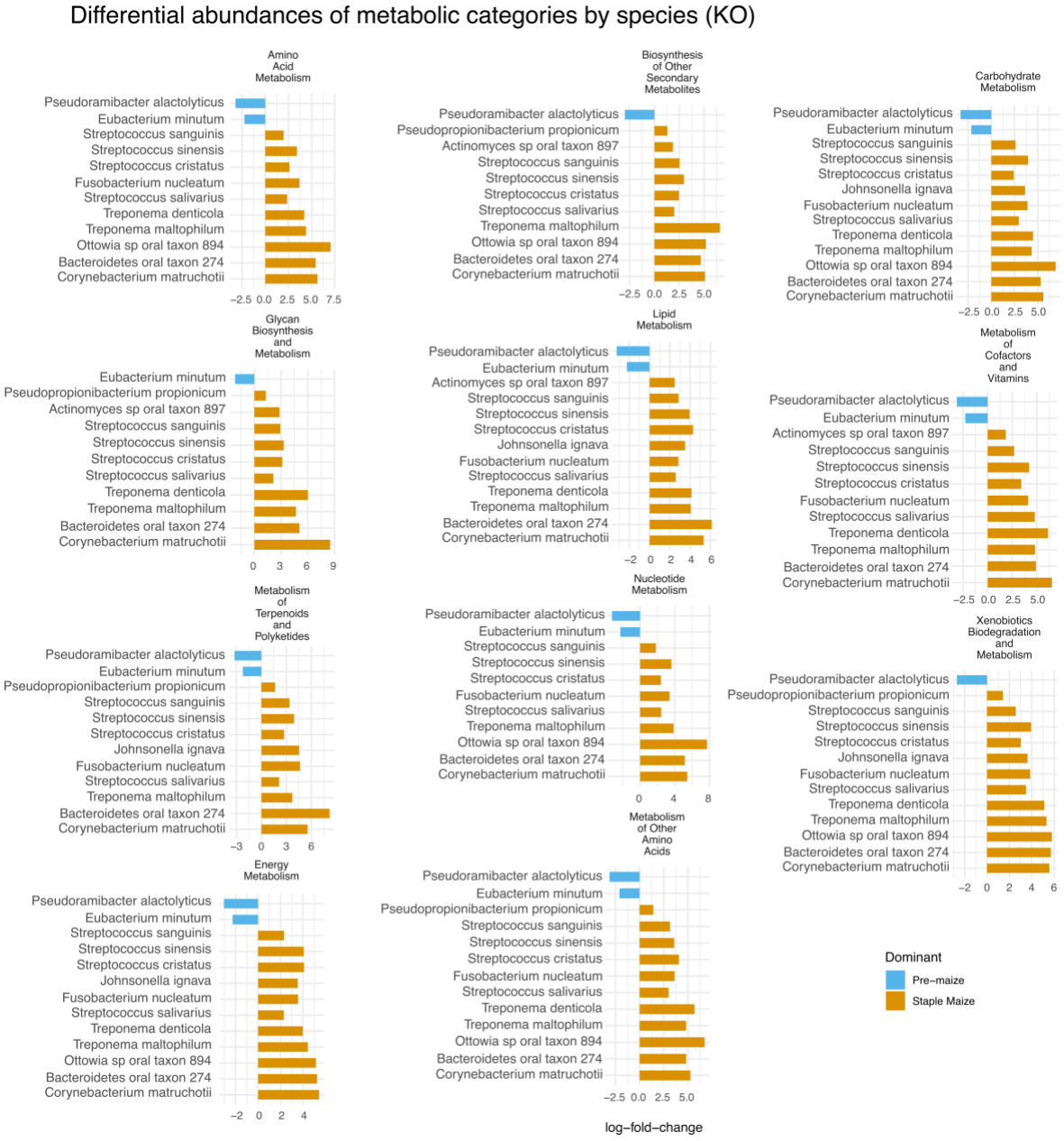


Figure 2.10 Taxonomic shifts across metabolic categories.

Log-fold-change of species determined by KO value with significant differences across subsistence categories. Pairwise Wilcoxon test performed with FDR correction.

Carbohydrate metabolic functions were further queried to observe log-fold-change across subsistence strategies and how the most influential taxa contribute to this metabolic category. Figure 2.11 demonstrates the individual species contributions to

different facets of carbohydrate metabolism across the SM and PM metagenomes. This figure is limited to the top five taxa contributing to either positive or negative PC1 ordination (top loading taxa). In the PM metagenomes, the relative abundances of individual species contributions to metabolic functions were consistently greater than individual contributions in SM metagenomes. The three primary contributors are *P. alactolyticus*, Anaerolineaceae bacterium oral taxon 439, and *E. minutum*. There is a more even distribution of all five top taxa contributing to carbohydrate metabolism in SM metagenomes, though the individual abundances were lower. When observing differential abundance of functional groups at a higher resolution (Figure 2.12), the KO groups with the largest observed difference in favor of SM metagenomes are related to lactose degradation (lactose PTS system EIIA component [EC:2.7.1.207], the pentose phosphate pathway, and the production of acetoin. Those more abundant in PM metagenomes are involved in acetate and formate pathways.

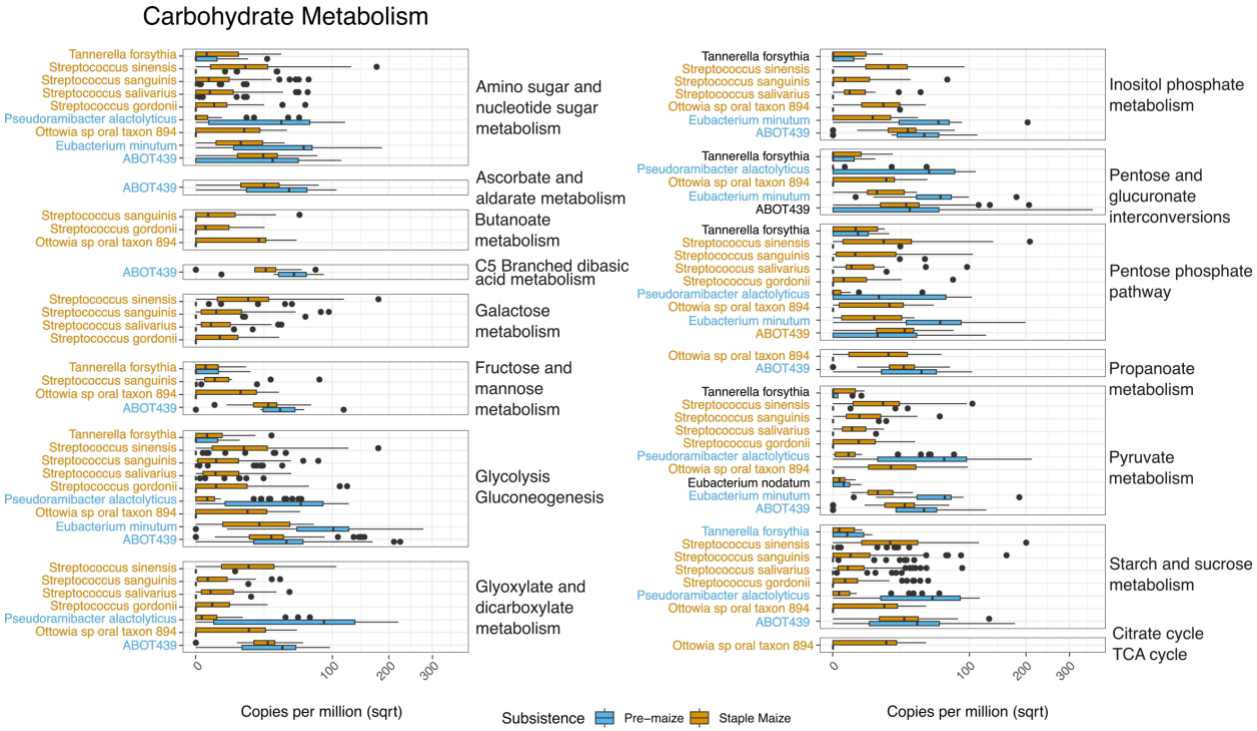


Figure 2.11 Carbohydrate metabolism boxplots.

Functional composition limited to most influential species identified by PCA bi-plot. (ABOT439 = Anaerolineaceae bacterium oral taxon 439). Species colored based on which subsistence pattern has the greatest abundance.

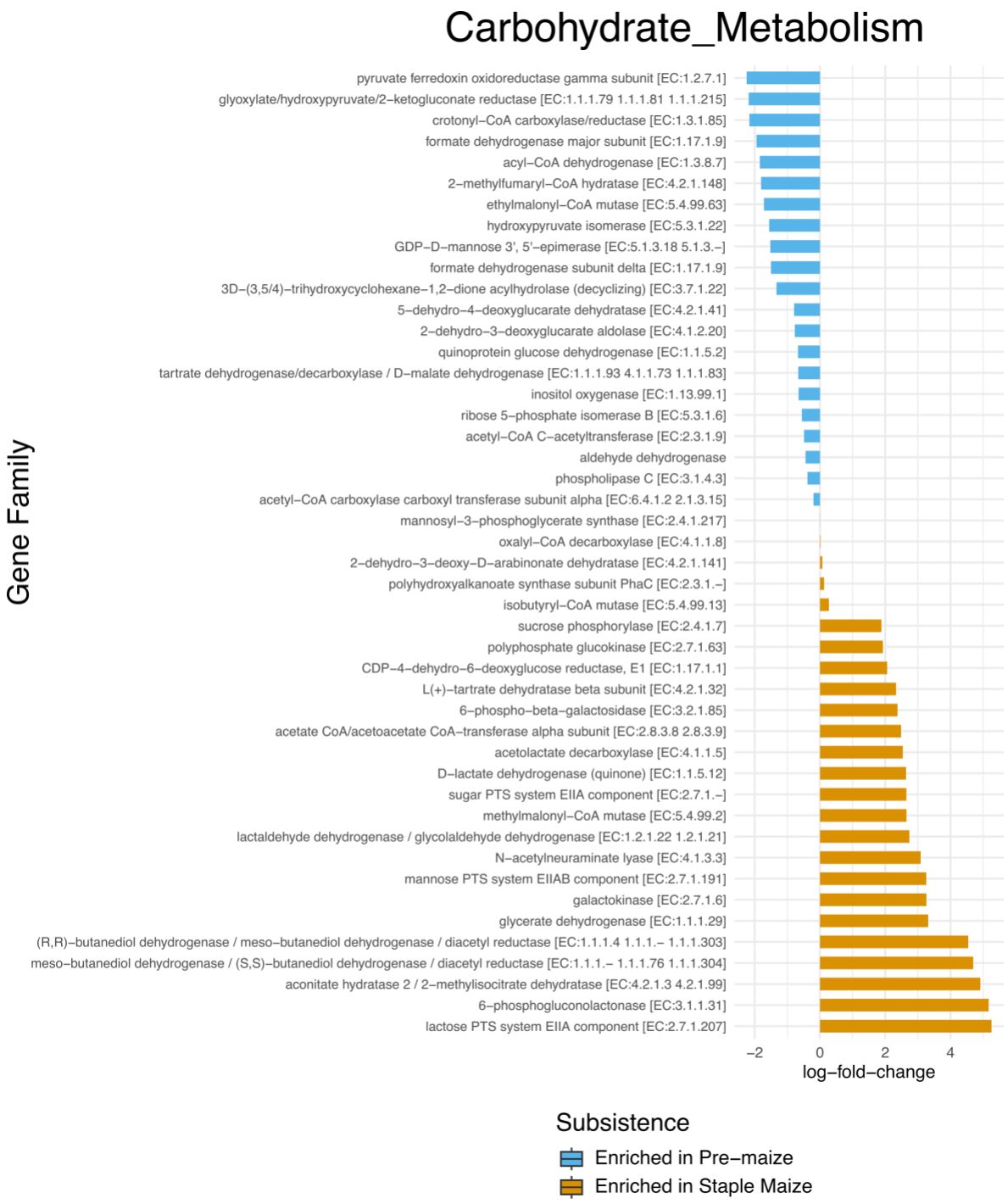


Figure 2.12 Functional shifts at the gene family (KO) level.

Only significantly differential (LFC, *p*-value < 0.05) functions within carbohydrate metabolism displayed.

We observed a consistent patterning in the taxonomic distribution across functional categories and subsistence strategies, necessitating further investigation. Specifically, we found that *Streptococcus* species dominate both the taxonomic and functional profiles of the SM metagenomes. To investigate why *Streptococcus* species, with their fermentative metabolism (Whiley & Hardie, 2015), replaced *P. alactolyticus* and *E. minutum*. As *Streptococcus* is aciduric, or capable of surviving at low pH, we categorized the functional profiles based on the ability for oral microbes to withstand low pH from saccharolytic fermentative reactions through acid tolerance and resistance (ATR) mechanisms.

2.4.4 Acid Tolerance and Resistance

To investigate mechanisms of acid tolerance and resistance, we parsed through gene family annotations for functions related to decarboxylation, malolactic fermentation reactions, urease and other deaminases, proton efflux pumps and cation antiporters, acid stress, cell membrane and cell density modification, chaperone production, and biofilm formation (Cotter & Hill, 2003; Krulwich et al., 2011; Lund et al., 2020). We examined annotations through both non-redundant UniRef50 gene families and the higher order KO groups to capture available ATR functional diversity.

ATR mechanisms were observed in Ancestral Maya metagenomes across the adoption of maize agriculture. The primary mode of pH regulation identified in PM metagenomes through UniRef50 annotations was through chloride-channel ion pumps and the predominant ATR mechanisms in SM metagenomes were those related to

molecular chaperones and biofilm formation. The taxa contributing to each of these functions annotated by the UniRef50 database are outlined in Figure 2.13.

On the order of KO functional groups, SM metagenomes are enriched in decarboxylase regulation and electrogenic transport. The log-fold-change of taxa contributing to KO ATR mechanisms across subsistence categories indicated that 19 species contributed to pH regulation across eight different mechanisms in SM metagenomes compared to two species across three mechanisms in PM metagenomes.

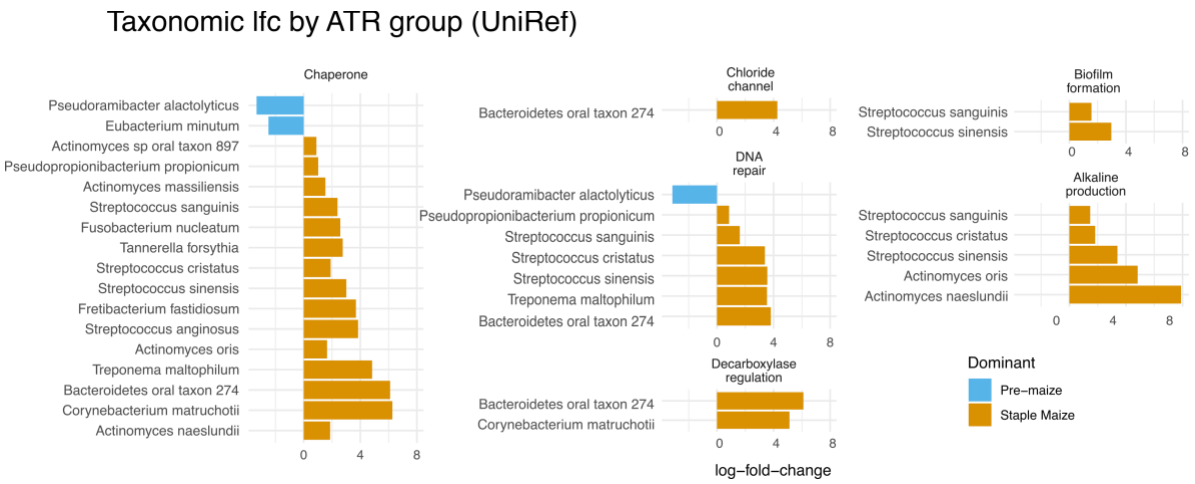


Figure 2.13 Taxa contributing to ATR (UniRef50).

Significant log-fold-changes in taxa contributing to different ATR functions (FDR adjusted *p*-value < 0.05).

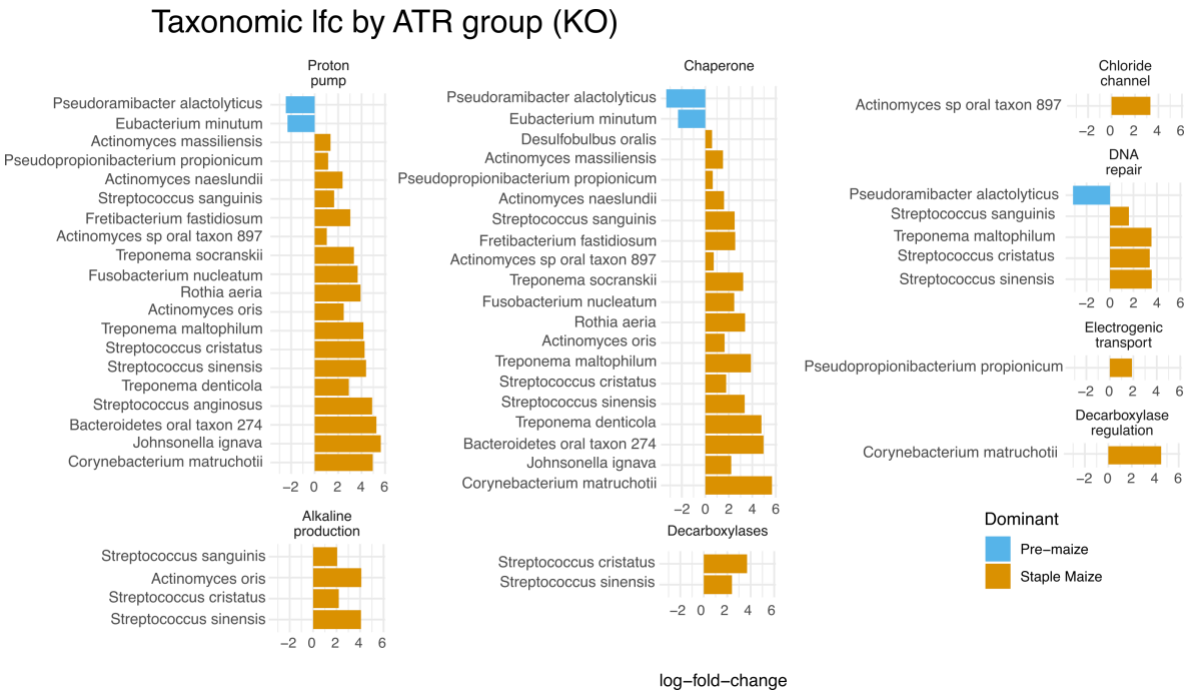


Figure 2.14 Taxa contributing to ATR (KO).

Significant log-fold-changes in taxa contributing to different ATR functions (FDR adjusted *p*-value < 0.05).

2.4.5 Genomic Reconstruction and Phylogeny

Ancestral Maya oral metagenomes were aligned to 16 of the most abundant oral microbes through a reference-based mapping approach to both authenticate the metagenomes as ancient and determine candidacy for genome reconstruction. After filtering alignment summaries for coverage and low heterozygosity ratios, we recovered genomes for *P. alactolyticus*, *Tannerella forsythia*, and *E. minutum*.

P. alactolyticus is not well characterized but has been implicated in endodontic infections (Siqueira & Rôças, 2003). The relative abundance of *P. alactolyticus* prior to the adoption of maize as a dietary staple distinguished these metagenomes from those after

the agricultural transition, in both taxonomic and functional composition. Genomes were only recovered from PM metagenomes (MRSC-1894, MRSC-2296) and the transitional metagenome (MRSC-1895) as too few reads from SM metagenomes mapped to the reference.

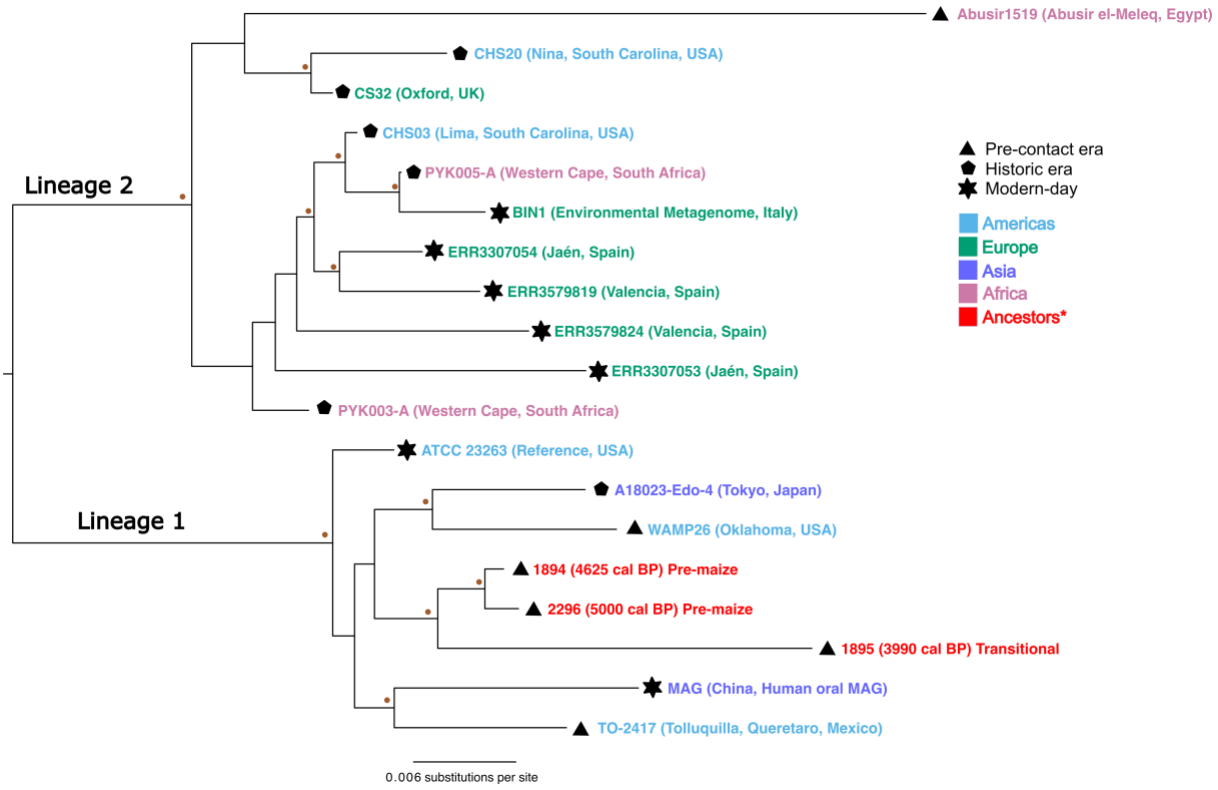


Figure 2.15 Maximum Likelihood tree for *Pseudoramibacter alactolyticus*.

The tree is based on an alignment of 7,985 homozygous bi-allelic SNPs, allowing upto 20% of missing data. The tree was built using the TVM+F+ASC+R2 model in IQ-TREE and rooted using a strain recovered from a wild-born chimpanzee. Brown circles indicate greater than 95% bootstrap support estimated using 1,000 ultrafast bootstrap replicates. Sample names are colored by geographic origin and accompanied by a symbol indicating time period.

As previously described (Fleskes et al., 2024), phylogenetic assessment of *P. alactolyticus* revealed two distinct lineages (Figure 2.15). Lineage 1 includes the reference strain (ATCC23263), as well as a modern metagenomically assembled genome (MAG) from

China and strains from historic Edo era, Japan, pre-European contact Mexico and Oklahoma, and all three Ancestral Maya strains. Lineage 2 includes all other modern genomes from Europe, historic South Africa, England, and South Carolina, and a modern environmental strain.

T. forsythia genomes were reconstructed from dental calculus of two SM individuals (MRSC-2060 and MRSC-2072) and five other individuals of mixed subsistence strategy after sequence capture (Figure 2.16). Tree topology for the *T. forsythia* phylogeny reinforces that described previously (Honap et al., 2023), with pre-European contact American strains in Lineage 1 and all other ancient, historic, and modern strains in Lineage 2.

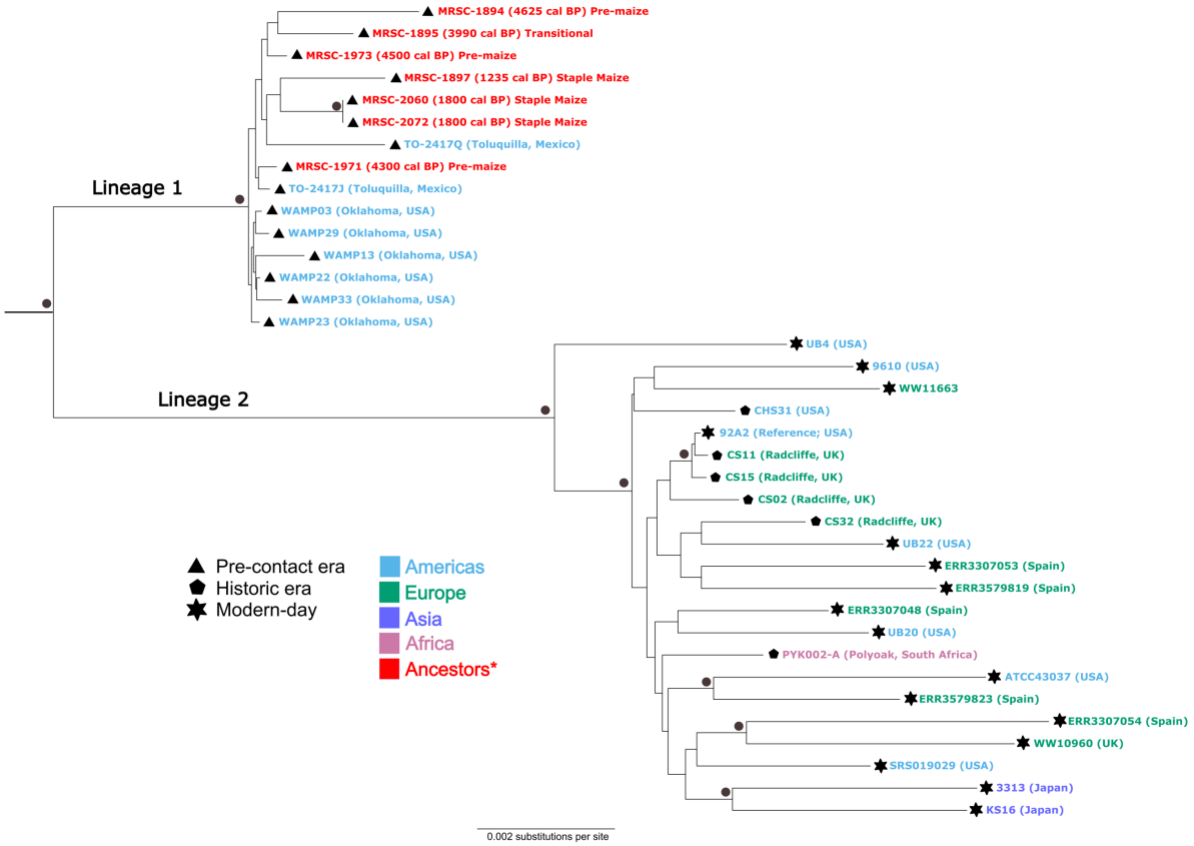


Figure 2.16 Maximum Likelihood tree for human *Tannerella forsythia* strains.

The tree was based on an alignment of 20,636 homozygous, bi-allelic SNPs, allowing for a maximum of 10% missing data. The tree was built using the TPM2u+F+ASC+R2 model in IQ-TREE and rooted using canine *T. forsythia* strains. Brown circles indicate branches with greater than 95% bootstrap support estimated using 1,000 ultrafast bootstrap replicates. Sample names are colored by geographic origin and accompanied by a symbol indicating time period.

Alignment to *S. sinensis* and *S. sanguinis* resulted in poor coverage at a depth of 5X and the heterozygosity ratio was almost exclusively greater than 1, suggesting the presence of multiple strains. Alignment to *E. minutum* resulted in eight genomes from PM and SM individuals covering more than 77% of the reference at least 5X with heterozygosity ratios between 0.03 and 0.37 suggesting a low likelihood of multiple strains. However, the bi-allelic SNP tree built with IQ-TREE did not have temporal or geographic structure Supplementary Figure 3. Future research characterizing essential genes for an *E. minutum* gene tree may help elucidate this poor resolution.

2.5 Discussion

Maize today is economically significant in Mesoamerica, providing the majority of calories from cereals and contributing to food security (Nuss & Tanumihardjo, 2010; Shiferaw et al., 2011). Its domestication from wild teosinte occurred in what is now southwestern Mexico before 9,000 cal. BP (Yang et al., 2023). Foraging populations facilitated the dispersion of semi-domesticated maize into the lower neotropics of Mesoamerica by 7,000 cal B.P (Kennett et al., 2022). Cultivation of maize was evident in the Maya lowlands by 6,500 cal B.P. and its intensification occurred after 5,000 cal B.P. By 4,000 cal B.P., maize was a dietary staple, constituting about 30% of total diet (Kennett et

al., 2020; Ray et al., 2024). The cultural shift towards agricultural production, experienced by other populations across the globe, had lasting repercussions for social and political structures, as well as overall health outcomes. Particularly, the transition irrevocably altered the oral microbial landscape. Human populations altered their surrounding environments through the intensification of agricultural production. They selected for singular vegetation in some areas and altered the distribution of nutrient sources for themselves and other species at different trophic levels. Cereal grains, like maize, are high in carbohydrates, altering the macronutrient proportion that was more varied for ancestral hunters and gatherers. Dietary carbohydrates promote the growth of saccharolytic bacteria, like *Streptococcus*, *Lactobacillus*, and *Actinomyces*, which metabolize sugars and produce acidic byproducts, reducing the pH of the oral cavity (Kianoush et al., 2014). Members of a healthy microbial flora facilitate the neutralization of environmental acidity, but the presence of rapidly digestible starches and fermentable carbohydrates, characteristic of refined grains and ultra processed foods, results in a greater relative abundance of cariogenic and acidogenic, or acid-producing, bacteria, increasing the risk of oral disease (Cohen & Armelagos, 1984; Cordain et al., 2005; Halvorsrud et al., 2019; Lommi et al., 2022; Mummert et al., 2011; Willis & Gabaldón, 2020).

The individuals presented in this study, who were consistently buried in a single geographic location over the span of about 7,000 years and across a major dietary transition, offer the unprecedented opportunity to observe oral ecological dynamics in response to a shift in host lifestyle. The adoption of maize as a dietary staple was likely accompanied by a transition in the proportions and types of macronutrients. The

proportion of carbohydrates in teosinte, the precursor for maize, is 70%, mainly in the form of starch with low amounts of free glucose and fructose (Bastias-Montes et al., 2020). Teosinte is high in leucine, glutamic acid, proline, tyrosine, phenylalanine, and arginine but has minimal aspartic acid, valine, lysine, isoleucine, alanine, glycine, and serine and lacks methionine and cysteine. Fatty acid composition is dominated by oleic, linoleic, and palmitic acid. Nutrient availability in the Maya Mountains was dispersed and seasonally varied so was not able to sustain large stable population structures prior to agriculture (Kennett et al., 2020). Isotopic values indicate that after the adoption of maize as a dietary staple, more than half of the individuals' dietary carbon was sourced from C_4 plants or animals eating such plants, which suggests stability in nutrient availability. Intuitively, we expect a consistent source for carbohydrates with respect to the SM individuals, likely contributing to a greater dietary proportion of starch.

Our analysis suggests that the dietary shift to stable maize consumption altered the taxonomic and functional composition of Ancestral Maya oral microbiomes. *Streptococcus* species largely replaced Clostridiales species, such as *Pseudoramibacter alactolyticus* and *Eubacterium minutum,* following this transition. The increase in intra-species diversity after the agricultural transition is antithetical to modern observations of salivary microbial responses to carbohydrate or sugar intake (Esberg et al., 2020; Millen et al., 2022) where alpha diversity is negatively correlated to increased consumption of carbohydrates. However, this observation has many caveats, many of which were addressed (Figure 2.6). Though the difference in phylogenetic diversity across the agricultural transition is significant, the same is not observed with respect to mean GC

content, rate of classification with MetaPhlAn4, or the number or average length of reads queried for taxonomic classification. Some other considerations must be accounted for. The disparity may be attributed to increased calculus formation from saccharolytic microbes effectively trapping more taxa throughout the individual's life. Alternatively, bacteria requiring a mixed community for growth are also differentially abundant in SM metagenomes, including *Actinomyces*, their symbionts of the candidate Saccharibacteria clade, and Bacteroidetes (Figueroa-Gonzalez et al., 2020; Naud et al., 2023), possibly contributing to the observed increase in species diversity. A likely cause for this observed increase in diversity may stem from either microbial preservation bias or an incomplete oral microbiome database. Archaeological samples have shown a loss of short, AT-rich fragments during sequencing preparation, though it is unclear if this is due to taphonomic processes or a product of double-stranded library preparation (Mann et al., 2018). Single-stranded preparation protocols could improve recovery of low-GC microbes, providing clarity to the question of taphonomic or methodological etiology of taxonomic skew, but alpha diversity may still be impacted if differential preservation is observed. Alpha diversity is strongly correlated with mean GC content (R=-0.95, p<0.001, Pearson Correlation) and classification rate (R=0.64, p = 0.001, Pearson Correlation), however as stated previously, these values do not significantly differ between subsistence groups and thus do not contribute to the observed difference in alpha diversity. The significant observation was not due to the reduction of metagenomic tables to oral taxa as this pattern is observed with and without filtering. However, the average rate of classification does not exceed 77% suggesting the possibility of missing data obfuscating accurate alpha diversity

measurements. About one third of the modern oral microbiome is uncultivable or not represented in available databases (Vartoukian et al., 2016). Previous attempts at metagenomically assembling oral microbes from plaque and tongue dorsa have revealed hidden microbial diversity in the mouth, particularly with respect to epibiotic bacteria, such as Saccharibacteria, Abscondidabacteria, and Gracilibacteria (Shaiber et al., 2020). This diversity was discovered in modern oral environments, which suggests that even further microbial diversity can be identified within the mouths of human ancestors. Though the observed differences in alpha diversity in the present study may be due to an impact of general subsistence pattern on the oral microbial profile, it is also possible that this is a result of either unequal preservation of microbial species through time or a product of an incomplete taxonomic database (Mann et al., 2018).

The pattern of taxonomic shifts is represented in the functional analyses. In terms of carbohydrate metabolic functions, PM metagenomes are enriched in acetate and formate pathways, both of which are end products of the highly abundant *P. alactolyticus* and *E. minutum* (Poco et al., 1996; Willems & Collins, 2015). SM metagenomes harbor a significantly greater abundance of functions related to lactose degradation, the pentose phosphate pathway, and the production of acetoin, a byproduct of fermentation by *Streptococcus* species (Hillman et al., 1987). *P. alactolyticus* is an obligate anaerobe, the growth of which is stimulated by fermentable carbohydrates (Willems & Collins, 2015). *E. minutum* is an asaccharolytic, or non-fermentative, obligate anaerobe with butyrate as its sole end product (Poco et al., 1996). Teosinte is high in resistant starches (Bastias-Montes et al., 2020) which are inaccessible to the aforementioned Clostridiales. Furthermore, *P.*

alactolyticus can grow in a limited pH range (pH=5.0-6.0) (Kianoush et al., 2014). Alternatively, Streptococci are facultatively anaerobic (Whiley & Hardie, 2015), readily ferment a variety of sugars (Abranches et al., 2018; Woo et al., 2002), and are capable of binding host amylase for the utilization of dietary starch (Fellows Yates (B) et al., 2021). A diet rich in carbohydrates, such as that of the SM individuals, promotes the growth of some acidogenic (acid-producing) or aciduric (acid tolerant) species associated with pathogenicity. *Streptococcus* species, as well as *Leptotrichia*, *Prevotella*, and candidate division TM7, are resistant to changes in pH (Abranches et al., 2018; Kianoush et al., 2014).

The production of acidic byproducts from the fermentation of carbohydrates, such as acetic, lactic, and propionic acid, degrade hydroxyapatite in enamel and dentine and contribute to the accumulation of carious lesions (Cotter & Hill, 2003; Kianoush et al., 2014). Those taxa resistant to fluctuations in pH, as well as some not included in this dataset such as *Scardovia wiggsiae*, *Parascardovia denticolens*, and *Lactobacillus salivarus* (Abranches et al., 2018; Marsh, 1994), have developed techniques to regulate pH. These include the consumption of protons through decarboxylase reactions, the production of ammonia from amine-containing amino acids, and the removal of protons through efflux pumps to lower pH in the cytoplasm. They have also evolved to evade the effects of acid stress by decreasing membrane permeability, increasing the production of chaperone systems, increasing expression of DNA repair mechanisms, and increasing biofilm production (Cotter & Hill, 2003; Lund et al., 2020). In this study, the primary mode of pH regulation in PM metagenomes identified through UniRef50 annotations was observed to be chloride-channel ion pumps, which prevent hyperpolarization of the inner

membrane after decarboxylation reactions (Iyer et al., 2002). This function is predominantly found in *E. minutum*, *P. alactolyticus*, and *Pseudopropionibacterium propionicum* (Supplementary Figure 1). The primary ATR mechanisms in SM metagenomes are those related to molecular chaperones and biofilm formation. Biofilm formation protects encapsulated microbes from the toxic effects of pH (Cotter & Hill, 2003), and these functions are found almost exclusively in *Streptococcus* and *Neisseria* sp. (Supplementary Figure 1). Molecular chaperones are beneficial in an acid-shock response to ensure proper functioning of DNA repair mechanisms (Cotter & Hill, 2003). These functions are identified in *Methanobrevibacter oralis*, Anaerolineaceae bacterium oral taxon 439, *E. minutum*, various *Streptococcus* species, *P. alactolyticus*, *Actinomyces* sp. oral taxon 897, *P. propionicum*, and *Propionibacterium* sp. oral taxon 192. The ATR mechanisms identified through KO annotation to be enriched in SM metagenomes includes decarboxylase regulation and electrogenic transport, both of which serve to regulate pH by consuming protons and expelling decarboxylation products (Foster, 2004; Iyer et al., 2002).

The alteration of their built environment and subsequent shift in nutritional availability for oral microbes likely drove the observed compositional differences. Oral microbiomes of SM individuals harbor a significantly greater abundance and variety of carbohydrate metabolic functions, primarily sourced from *Streptococcus* species. We also found that across the dietary transition, the number of ATR mechanisms increased, as well as the diversity of taxa contributing to such functions, indicating that environmental pH may have contributed to the compositional shift in oral microbial taxonomy and

function. Further research could explore other environmental states that accompanied agricultural transitions that may have shaped oral microbial ecologies, such as food processing, cooking, storage, or texture. Additional exploration could also address the nutrient contributions of food items complimentary to maize agriculture, such as squash or beans.

The sequential deposition and subsequent calcification of microbial plaque biofilm over the course of an individual's life encapsulates plant microremains, such as phytoliths and starch granules, that have been used to reconstruct diet and infer environmental conditions in ancient human populations (Modi et al., 2020, 2023; Quagliariello et al., 2022). The presence of such material suggests that DNA from its source may also be trapped within the calcified matrix. However, investigation into this assumption has revealed that positive identification of plant DNA requires sufficient read depth and coverage to validate that the plant DNA is ancient and that eukaryotic identification is not due to read mis-mapping (Brealey et al., 2020; Mann et al., 2023; Moraitou et al., 2022). Though we were able to detect a maize signature with the Burrows Wheeler aligner (Li & Durbin, 2009), we were unable to detect dietary DNA through screening with a reduced database. Thus, large-scale dietary discovery in dental calculus, something that has historically been difficult (Brealey et al., 2020; Mann et al., 2023; Moraitou et al., 2022; Ottoni et al., 2021; Weyrich et al., 2017), would need an improved approach. Furthermore, the metagenome with the most uniquely aligned reads (> 8000) was the TM individual rather than SM. This is likely due to the quantity of AR reads included in the analysis. The number of unique reads mapping the *Z. mays* genome is strongly correlated to the number

of AR reads queried (Supplementary Figure 4), of which the TM calculus metagenome has the greatest amount. Evidently, the consistent identification of direct dietary constituents remains elusive.

Differential abundance of microbes across host lifestyle transitions are recapitulated in phylogenetic reconstructions of these taxa. *P. alactolyticus* genomes recovered from Ancestral Maya share a phylogenetic lineage with modern and historic eastern Asian strains, suggesting that it accompanied humans through initial continental expansion into the Americas. Additionally, the modern American reference falls within this lineage (Lineage 1) indicating this strain shows persistence through modern time. The only American strains falling in the alternative lineage, including all ancient, historic, and modern strains from Africa and Europe, were isolated from African-descended individuals from 18th century Charleston, South Carolina (Fleskes et al., 2024). Though these individuals were born in the US, they retained the Lineage 2 *P. alactolyticus* strain either from generations of caregivers through vertical transmission, or through horizontal community acquisition via other African-descended individuals or European-descended slavers. It is evident from this population that the transition to agriculture reduced the relative abundance of this organism, perhaps due to its vulnerability to low pH or its weakness in the face of competition from other microbes. It is not yet clear what has driven this phylogenetic structuring, but its response to shifting dietary practices may play a role. More widespread identification of this bacterium may illuminate more information.

T. forsythia is an opportunistic pathogen implicated in periodontal disease alongside *Treponema denticola* and *Porphyromonas gingivalis* of the red complex

(Socransky et al., 1998; Socransky & Haffajee, 2005). It is indicated by PCA to affect sample grouping in positive ordinal space, though its relative abundance does not significantly shift between subsistence patterns. It's contribution to metabolic function varies across dietary practices and only significantly increases for peptidoglycan biosynthesis (Supplementary Figure 2) and the ATR mechanism of chaperone function (Figure 2.13). The tree topology for *T. forsythia* suggests an alternative evolutionary history to that seen with *P. alactolyticus*. The strains identified in Ancestral Maya metagenomes, from both PM and SM individuals, cluster exclusively with other strains from pre-European contact America (Lineage 1). All ancient or historic European, Asian, or African strains cluster with all modern strains globally in Lineage 2, indicating that the American *T. forsythia* strain was replaced by that introduced after contact. With no apparent structuring corresponding to subsistence strategy, dietary practice may not play a significant role in its evolutionary response. The resolution of this history could be more refined with further *T. forsythia* genome reconstructions across different time periods and geographic locations.

2.6 Conclusions

This study represents the first investigation into the impact of an agricultural transition on oral microbial composition and function in the Americas. Continuous deposition of human remains at two rock shelters in the Maya Mountains of Belize and isotopic data contextualizing maize consumption provides an opportunity to explore microbial response to a shifting subsistence strategy without the confounding effects of

geography. The adoption of maize as a dietary staple in Mesoamerica approximately 5,000 cal. BP was accompanied by a shift in microbial structure. After that time, maize was the primary source of carbohydrates and starch as reflected in dietary isotopes of carbon from collagen and bone apatite (Ray et al., 2024). Clostridiales bacteria, including *Pseudoramibacter alactolyticus* and *Eubacterium minutum*, significantly decreased in both relative abundance and in contributions to metabolic functions. The macronutrient composition of a maize-based diet favored the proliferation of saccharolytic bacteria, including *Streptococcus* sp., *Actinomyces* sp. and their intracellular symbionts, Candidatus Saccharibacteria. This pattern of replacement of Clostridiales bacteria by *Streptococcus* and other species may be attributed to their capacity for pH regulation, as the degradation of starch and the fermentation of sugars produces acidic by-products. Mechanisms of acid tolerance and resistance are not universally distributed among oral microorganisms but are found in greater abundance and diversity among taxa in the staple maize metagenomes. Though the adoption of maize corresponds to differential abundances of some taxa and a shift in metabolic composition, the evolutionary mechanisms driving bacterial strain diversity are not so discernable. The significant decrease in abundance of *P. alactolyticus* following staple maize consumption points to a dietary influence on bacterial survival, however the American *P. alactolyticus* strain found in the Ancestral Maya has persisted till today despite the influx of strains from Africa and Europe following colonization and further changes in diet post-industrialization. Conversely, *Tannerella forsythia* strains recovered from the Ancestral Maya are similar to those found in other pre-European contact American populations, however, this lineage

seems to have been replaced following colonization, suggesting an alternative driver of strain vulnerability and persistence. The bacterial network dynamics and evolutionary phenomena presented here reveals a level of complexity in the interchange between health, diet, and microbial ecology across human evolutionary landscapes.

CHAPTER 3 – ORAL METAGENOMES OF MUWEKMA OHLONE ANCESTORS IN CALIFORNIA OFFER INSIGHTS INTO PRE-COLUMBIAN MICROBIAL STRAIN DIVERSITY AND ECOLOGICAL DYNAMICS

3.1 Abstract

Prior to European colonization, people living in the coastal Pacific region practiced hunting and gathering for millennia and subsisted on acorns and seeds as a dietary staple. Their oral microbiomes offer the opportunity to observe microbial community dynamics in a pre-agricultural population eating a diet high in starch. Here, in collaboration with the Muwekma Ohlone Tribe of the San Francisco Bay Area, we studied the oral microbiomes of 69 Ancestors from 13 archaeological sites spanning the Early (EP; ~3,670 YBP; n=2), Middle to Late (MP; ~2,450 YBP-European contact; n=64), and Mission (SP; C.E. 1780s-1790s; n=3) Periods. DNA was recovered from dental calculus, partially treated with uracil-DNA-glycosylase, built into double-stranded libraries, and shotgun-sequenced with Illumina technology. Taxonomic and functional composition was then assigned to the resulting metagenomes. Metagenomically assembled genomes (MAGs) of microbes were obtained through *de novo* assembly of Muwekma Ohlone and other pre-European contact American metagenomes. The Ancestors' microbiomes were enriched in Actinobacteria, Firmicutes, and Synergistaceae bacteria. Functions related to starch and sucrose metabolism, such as those involving oligosaccharide metabolism and biofilm formation, were elevated. *De novo* assembly resulted in the identification of 132 representative MAGs, which includes the first *Streptococcus sinensis* genome recovered from archaeological contexts.

Phylogenetic analysis of MAGs through alignments of core genes and genome-wide single nucleotide polymorphisms (SNPs) helped elucidate evolutionary relationships of three oral taxa. This study presents the largest ancient oral metagenomic dataset from the Americas thus far and reveals the oral microbial composition and hidden strain diversity in a population with a unique diet.

3.2 Introduction

Human populations operate within ecological networks, adapting to their environments by leveraging the surrounding resources. People inhabit the land and interact with the local fauna, competing for resources, fragmenting habitats, hunting the birds and mammals, domesticating livestock, fishing, and collecting coastal marine life. They modify the environment through building settlements, land management or agricultural practices, and collecting and storing local vegetation. This interconnectivity at on a macroscopic scale is impacted by, and reciprocally impacts, microorganisms associated with each element of the network.

This connectivity firmly places microbiome research within a one health framework, which states that the health of any one of the three systems of life, humans, animals, and the environment, is contingent upon the health of the other two (Huang et al., 2024; Silbergeld, 2019). A glimpse into the history of microbiomes that accompany human populations is often limited, not just by the extent to which biological material is preserved through time, but also by our ability to fully describe microbial communities with any degree of nuance. The exploration of host-associated microbial communities in antiquity

has proliferated within the context of the oral microbiome. We directly engage within the one health system through our mouths with the food we eat, the medications we consume, the pathogens we encounter, and oral hygiene practices. Oral-associated microbiomes are exposed to these interactions and can thus reflect the dynamics between the host and surrounding ecosystem. Plaque biofilms, or dental calculus, calcify over the life course of an individual, trapping and preserving microbial genetic information for centuries. However, the exploration of oral microbial diversity has been disproportionately localized in the eastern hemisphere, and the individual microorganisms highlighted in evolutionary investigations are often of modern clinical significance. Few studies have contextualized oral microbial evolution within the context of the last great continental expansion of humans into the Americas.

When human groups first populated the American continent, they were accompanied by their oral microbiota. People migrated across the continents, settled in diverse climatic regions and leveraged the resources of their surrounding environments. For the coastal regions of the eastern Pacific, this included pinnipeds, shellfish, terrestrial herbivores, and the acorns and seeds from tanoaks and coastal live oaks that were ubiquitous throughout the region (Milliken et al., 2007; Wohlgemuth, 1996). Coastal Pacific populations in the Early Period (EP; 5450-2450 YBP) primarily relied on large-bodied marine mammals, such as the northern fur seal. There is also some evidence of acorn leaching, meaning acorns were processed to remove toxic tannins before consumption (Milliken et al., 2007). Through the Middle and Late Periods (MP; 2,450 YBP-European contact), there was a shift towards greater settlement permanence and increased variation in marine and

land herbivore food sources (Monroe, 2014). Notably, this period is marked by intensification of acorn consumption to a dietary staple and its rendering as a flour (DiGiuseppe et al., 2021; Milliken et al., 2007). These populations actively participated in land modification through prescribed burns to promote the growth of these small-seeded plant resources (Lightfoot et al., 2013; Wohlgemuth, 1996). The arrival of the Spanish settlements in the 18th century led to the establishment of missions, which significantly altered the lives of the native peoples. Indigenous coastal pacific groups were thrust into agricultural crop production in the Mission Period (SP; 18th century), diverging from ancestral hunting and gathering practices and prevented from traditional land management (Leventhal et al., 2011; L. M. Panich et al., 2024). This study seeks to explore the impact of these varying subsistence practices and colonial imperialism on microbial ecological dynamics encapsulated in the calcified matrix of dental calculus.

Calculus is not only an ideal medium to investigate microbial composition and function in the context of dietary practices but is also an excellent source to study the evolutionary trajectory of opportunistic pathogens implicated in oral disease and commensal organisms of less clinical significance but great relevance to the ecological dynamics of an oral microbial community. The sequencing of DNA harbored in ancient dental calculus results in relatively short fragment lengths due to its pattern of degradation (Dabney (B) et al., 2013). The nature of microbial genomic structures, such as insertions, deletions, repeat regions, or AT-rich regions, is not always conducive for the recovery of complete microbial genomes from short DNA sequencing reads (Mann et al., 2018). Traditional genomic recovery methods, such as through alignment to a published

reference genome, can limit the detection of divergent strains, unrecognized species, or gene-loss events (Granehäll et al., 2021). However, advancements in methods pertaining to *de novo* metagenome assembly have vastly expanded detectible microbial diversity by increasing the description of previously uncultivable taxa through the construction of metagenomically assembled genomes (MAGs) and the identification of species-level genome bins (SGBs) (Almeida et al., 2021; Pasolli et al., 2019). Since the successful recovery of MAGs through ancient metagenome assembly (Brealey et al., 2020), many studies have recapitulated this process to elucidate past microbial composition and function without an apparent effect of DNA damage on the quality of ancient metagenomic assembly (Standeven et al., 2024; Wibowo et al., 2021). *De novo* genome reconstruction from metagenomes has the potential to uncover hidden microbial diversity in the Americas over the past millennia.

In collaboration with the Muwekma Ohlone Tribe of the San Francisco Bay area and the Amah Mutsun Tribal Band of the Mission San Juan Bautista, we investigated oral microbial diversity of California Ancestors. The Muwekma Ohlone Tribe is the California Native American Heritage Commission's most likely descendant tribe for individuals from 12 of the 13 archaeological sites, including all that pre-dated European colonial expansion and subsided on local resources through hunting, gathering and fishing. Three individuals from the Sanchez Adobe lived during the Mission period and were thus agriculturalists. Here, we present metagenomes recovered from 76 individuals interred between ~3670 and ~240 years before present (YBP), making this the largest ancient oral microbiome dataset from a single geographic area in the Americas. We explore microbial composition and

function of these hunter-gatherer metagenomes with a unique diet of high-starch acorns and seeds. By contextualizing these oral metagenomes amongst others from various subsistence strategies, we highlight continuities and distinctions of the dietary impact on these oral microbiomes. Additionally, through metagenomic assembly and phylogenetic analysis, we investigate the effects of globalization and colonialism on the evolutionary trajectories of oral taxa.

3.3 Materials and Methods

3.3.1 Study Population

Muwekma Ohlone and Amah Mutsun Ancestors

Dental calculus was sampled from teeth of 75 Ancestors from 13 archaeological sites surrounding what is now the San Francisco Bay area. Two individuals recovered from the Horton Street landing project (CA-ALA-312) were dated to the Early Period (EP; 5450-2450 YBP). Most of the individuals (n=70) were recovered from 11 sites and dated to the Middle (2,4500-900 YBP) and Late (900 YBP-European contact) periods (MP). A sample was collected from the abscess of one of these individuals, resulting in 71 total samples from the MP. Three individuals recovered from the Sanchez Adobe in San Mateo County (CA-SMA-71) were dated to the Mission Period (SP; 1780s-1790s C.E.). Table 3.1outlines the archaeological sites including the number of samples recovered from each site, the shell bead period, and the California site names.

Site ID	Site Name	Number of samples	Temporal period

CA-ALA-312	Horton Street Landing Project	2	Early Period
CA-ALA-329	Mánni Muwékma Kúksú Hóowok Yatiš Túnneštе-tka	1	Middle-Late Period
CA-ALA-695	Van Daele	18	Middle Period
CA-SCL-128	Thámien Rúmmeytak	13	Middle Period
*CA-SCL-134	Kalawwasa Rummeytak	7	Upper Middle Period
CA-SCL-215	Yakmuy 'Ooyákma-tka	9	Middle Period
*CA-SCL-287	Yuki Kutsuimi Šaatoš Inūxw	7	Middle Period
CA-SCL-343	Lands of Oyama	2	Middle-Late Period
CA-SCL-851	’Utthin Širkeewis Tcitca ’Irekmatka	1	Middle Period
CA-SCL-867	Ríipin Waréeptak	2	Middle Period
CA-SCL-869	Katwáš Ketneyma Waréeptak	1	Middle Period
CA-SFR-191	Ralston Mound	9	Middle Period
CA-SMA-71/H	Sanchez Adobe	4	Mission Period

Table 3.1 Muwekma Ohlone and Amah Mutsun archaeological sites.

*Includes samples previously processed by co-author, J.L.

Individuals of the EP obtained much of their diet from animal protein, relying on large game from terrestrial and marine sources (Leventhal et al., 2017; Milliken et al., 2007). The MP is marked by a transition to smaller land animals, shellfish, and the intensification of acorn and seed consumption and no evidence of C_4 plant consumption (Gardner et al., 2018; Leventhal et al., 2010; Wohlgemuth, 1996). A stark division of labor between men and women throughout the MP has been hypothesized, with boys weaned at an earlier age and consuming a higher quantity of meat from game. Girls were weaned later and likely acquired protein from marine sources, such as shellfish (Eerkens et al., 2022). Individuals from the SP practiced an agricultural subsistence economy enforced by the Spanish colonizers.

Comparative populations

To contextualize the taxonomic composition and functional potential of the California metagenomes, comparative metagenomes from dental calculus were included

in some analyses, including pre-European contact agricultural Wichita Ancestors from what is now Oklahoma (Honap et al., 2023), Ancestral Maya in Belize from before and after the adoption of maize as a dietary staple (Chapter 2), historic era (18th-19th century) Europeans from the Radcliffe Infirmary at Oxford, and modern-day Spanish individuals (Fellows Yates (B) et al., 2021; Velsko et al., 2019).

Population	Author	Number of Samples	Publication
MRSC - Pre-maize	Johnson2025	14	Unpublished
MRSC - Staple Maize	Johnson2025	9	Unpublished
Wichita	Honap2023	26	10.1002/ajpa.24735
Radcliffe Infirmary	Velsko2019	43	10.1186/s40168-019-0717-3
Modern	Velsko2019	10	10.1186/s40168-019-0717-3
Modern	FellowsYates2021	8	10.1073/pnas.2021655118

Table 3.2 Comparative populations.
Details for each population used in the comparative analysis.

The pre-European contact American groups provide a range of subsistence strategies from three different geographic locations in the western hemisphere. The Pre-maize (PM) individuals from Belize predate the proliferation of agricultural practice in the Americas but offer a contrast to the hunting and gathering lifestyle practiced by the EP and MP Muwekma Ohlone Ancestors. The Staple Maize (SM) individuals in Belize and the Wichita Ancestors were agriculturalists with staple maize consumption, offering an example of high-starch consumption from a staple crop rather than the acorn and seed collection practiced by EP and MP Muwekma Ohlone Ancestors. Though these agriculturalists predate European contact, they offer a comparison to the SP individuals who were required to mirror the agricultural lifestyle of Spanish colonizers. The historic UK

and modern Spanish individuals pose as post-industrial counterparts to the comparative agricultural Wichita and SM individuals, further contextualizing the California Ancestors across a temporal continuum. The distinct characteristics of these populations help to highlight those aspects of the Muwekma Ohlone metagenomes associated with their unique diet.

3.3.2 Sample Preparation

All calculus sampling, DNA extraction, and library preparation was performed in the dedicated clean lab at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR). Calculus was removed from the surface of teeth, and when multiple teeth were present for an individual, calculus was combined. Detailed methods are described in Chapter 2 of this dissertation (Chapter 2). Briefly, calculus was subsampled to use 1-10 mg of material, when possible, and irradiated to remove surface contaminants. The 76 samples were processed for DNA extraction over six batches, each accompanied by an extraction control. Following enzymatic digestion with 0.5 M EDTA (pH = 8.0) and proteinase K over the course of seven days for each batch, DNA was extracted using a modified silica column-based method optimized for the recovery of short DNA fragments (Dabney (A) et al., 2013; Ozga et al., 2016). Extracts were partially treated with uracil-DNA-glycosylase (UDG) to repair most damage characteristic of aDNA while retaining nucleotide transitions at the 5` and 3` termini for ancient DNA authentication (Rohland et al., 2015). Double-stranded libraries were prepared using the blunt-end single tube (BEST) method (Carøe et al., 2017), with a library control included for each batch. Libraries were

dual-indexed over four replicate reactions to maximize library complexity and randomize biases occurring during PCR amplification. Library replicates were pooled, purified, and analyzed using the D1000 Assay on the Agilent TapeStation as given in (Honap et al., 2023). Libraries presented in this study were shotgun-sequenced across multiple paired-end runs on an Illumina HiSeq 2x150 at Admera Health in South Plainfield, NJ. A subset of libraries was sequenced on an Illumina NextSeq 2x150 at the Oklahoma Medical Research Facility (OMRF) in Oklahoma City, OK.

3.3.3 Data Preprocessing

Paired-end FASTQ files from this study and those for comparative datasets were processed for removal of Illumina adapters, ambiguous (N) bases, and low quality (PHRED < 30) bases were trimmed. Reads shorter than 30 nucleotides (nt) were excluded, and paired-end reads with an overlap of at least 10 nt were collapsed using AdapterRemoval v2.3.3 (Schubert et al., 2016). PCR duplicates were removed using fastp v0.23.4 (Chen, 2023).

Deduplicated, merged reads from this study were aligned to the human reference genome (T2T-CHM13v2) using Bowtie v2.5.1 (Langmead & Salzberg, 2012) with default parameters to determine endogenous host DNA content. Mapped and unmapped reads were sorted and separated with SAMtools v1.9 (Danecek et al., 2021). Unmapped reads were converted to FASTQ files using bam2FastQ from BamUtil v1.0.15 (Jun et al., 2015) and used for downstream analysis of microbial content; these FASTQ files are hereafter referred to as analysis ready (AR) reads. Reads mapped to the human genome were

deduplicated with DeDup v0.12.8 (Peltzer et al., 2016) and the frequency of terminal nucleotide transitions was assessed with MapDamage v2.2.1 (Jónsson et al., 2013).

3.3.4 Taxonomic Profiling

AR reads were used for community taxonomic analysis through a dual-pronged approach using MetaPhlAn4 and Kraken2. MetaPhlAn4 allows for prokaryotic community profiling using a mechanism and database similar to that of the gold-standard functional profiling tool, HUMAnN (Beghini et al., 2021), thereby allowing for direct comparison of community taxonomic and functional profiles. MetaPhlAn uses a clade-specific marker gene database of bacterial and archaeal taxa, resulting in fewer false positive hits among the taxa detected, albeit potentially reducing sensitivity of the profiling. On the other hand, Kraken2 allows for the use of custom databases comprising whole genomes as well as metagenomically assembled genomes (MAGs) of prokaryotes, eukaryotes, and viruses. While the use of whole genomes in the Kraken2 database increases the sensitivity of the profiling, it can also lead to more false positive hits, necessitating further authentication of taxa detected (Ottoni et al., 2019, 2021; Ozga & Ottoni, 2023).

Reads were classified with MetaPhlAn v4.0.2 (Blanco-Míguez et al., 2023) using the database version mpa_vJan21_CHOCOPhlAnSGB_202103, and a minimum read length of 30 nt. Reads were also screened by exact *k*-mer matching through Kraken v2.1.2 (Wood et al., 2019) against a custom database of representative bacteria and archaea from the Genome Taxonomy Database (GTDB) and fungi, protozoa, viruses, and the UniVec core RefSeq sequences, as well as plastids and mitochondria, from the National Center for

Biotechnology and Information (NCBI). The Kraken2 database was built with default parameters and the accompanying Bayesian re-estimation tool, Bracken v,2.0 (Lu et al., 2017), was built with a minimum read and *k*-mer length of 35 nt. For all community-level comparative analyses, taxa were filtered to only include oral-associated microbes to stringently avoid the influence of possible environmental contaminants. Taxa were identified as oral if they were included in the human oral microbiome database (HOMD) (Escapa et al., 2018), they were identified as contributing to the oral community after source tracking, or if the species-level genome bins (SGBs) were isolated from an oral source.

3.3.5 Source Tracking and Ancient Authentication

To ensure that all downstream analyses would not be influenced by environmental contaminants, source tracking was employed to measure the proportion of microbial profiles attributed to an oral source. Using previously published metagenomic profiles as "sources", including modern dental calculus (Velsko et al., 2019), supra- and subgingival plaque (HMP, 2012), feces from populations leading industrialized and non-industrialized lifestyles (Obregon-Tito et al., 2015; Rampelli et al., 2015), skin (Oh et al., 2014), and soil (Johnston et al., 2016), the proportion of contributing environments was assessed for each ancient metagenome, or "sink", using SourceTracker v2.0.1 (Knights et al., 2011). Metagenomes were rarefied to 30,000 reads and a prior species count of 0.01 relative to the number of sequences was added to training environments (alpha1) and 1.0 relative to the input was added to the unknown environment (alpha2) to prevent overfitting (Gancz et

al., 2023; Knights et al., 2011). The resulting tables were processed in R and plotted using ggplot2 v3.5.1 (Wickham, 2016).

To validate the antiquity of the metagenomes, AR reads were aligned to prevalent oral taxa to measure the frequency of cytosine-to-thymine (C-to-T) and guanine-to-adenine (G-to-A) changes at the terminal nucleotides. Reads were aligned to the five most abundant taxa identified by MetaPhlAn4 using BWA aln v0.7.17 (Li & Durbin, 2009) with seeding disabled and –n parameter set to 0.1 to enable a strict mapping approach. Mapped reads that were sorted and deduplicated with SAMtools (Danecek et al., 2021) and DeDup (Peltzer et al., 2016), respectively, were then evaluated with MapDamage (Jónsson et al., 2013). For each reference identified with greater than 600 uniquely aligned reads, nucleotide transitions were plotted in R to view the damage patterns of prevalent oral taxa.

Each metagenome with greater than 25% of taxa attributed to an oral source, evidence of damage diagnostic of aDNA, and at least 500,000 AR reads were retained for downstream analyses.

3.3.6 Functional Profiling

To assess the impact of Muwekma Ohlone and Amah Mutsun dietary practices on oral microbiota, metagenomes from this study and other comparative populations were functionally profiled using HUMAnN v3.6 (Beghini et al., 2021) using the UniRef50 search mode against the v201901_v31 ChocoPhlAn database and a minimum read length of 30 nt. Annotated gene families were re-normalized from reads per kilobase (RPK) to copies per million (CPM) to account for sequencing depth. Normalized tables were then regrouped

into KEGG orthologous (KO) groups (Kanehisa & Goto, 2000) to further classify functions into higher-order metabolic categories with the available KEGG modules (Kanehisa et al., 2014). The regrouped gene family table stratified by contributing taxa was filtered to only keep units with greater than 2 CPM in at least 20% of samples. A feature table mapping KO functions to higher order metabolic modules and pathways was obtained from HUMAnN legacy data (Abubucker et al., 2012) to further categorize functional composition.

3.3.7 Statistical comparisons

All statistical comparisons were performed in R v4.2.2 and data tables, taxonomic lineages, and sample metadata were stored in phyloseq objects (McMurdie & Holmes, 2013). For all comparative analyses, compositional data were centered log-ratio (CLR) transformed to remove individual variables from bound space and allow for any statistical method to be used (Quinn et al., 2019). To observe compositional dynamics with respect to sample metadata for California metagenomes, the data were ordinated in a principal component plot (PCA) using the mixOmics library v6.22 (Rohart et al., 2017), plotted with ggplot2 (Wickham, 2016), and a permutated analysis of variance (PERMANOVA) was performed on principal components using vegan v2.604 (Oksanen et al., 2001). The betadisper function from vegan followed by analysis of variance (ANOVA) was used to determine if any observed statistical significance was due to heterogeneity of groups dispersions rather than composition.

Comparative taxonomic and functional tables were analyzed for inter-group (beta) diversity by calculating Euclidean distance on CLR-transformed data and plotting with a

principal coordinate analysis (PCoA). Relative abundances of top species visualized with heatmaps plotted in ggplot2 (Wickham, 2016) using geom_tile. To determine the species and functions differentiating populations, a random forest machine learning algorithm was employed using the randomForest v4.7-1.2 (Liaw & Wiener, 2002) package. The random forest model was built using 500 trees and variable importance and proximity data were reported. The model was cross-validated with trainControl from the caret v7.0-1 package (Kuhn, 2008). The relative abundances of those variables deemed important by the models were plotted in ggplot2 (Wickham, 2016).

3.3.8 Assembly and MAG Construction

Individual metagenomes from this study, including blanks, and those of Wichita and Ancestral Maya, were *de novo* assembled following the pipeline outlined in (Wibowo et al., 2021). Briefly, metagenomes were assembled using MEGAHIT v1.2.9 (Li et al., 2015) with a minimum and maximum *k*-mer size of 21 and 91, respectively, and a *k*-step of 10. Resulting contigs were filtered to a minimum of 1000 nt using SeqKit (Shen et al., 2016) and short reads were mapped back to the filtered contigs using Bowtie v 2.3.5.1 (Langmead & Salzberg, 2012) with the --very-sensitive preset and -N 1 to allow for additional mismatches in ancient DNA. Alignments were sorted and indexed using SAMtools v1.9 (Danecek et al., 2021) and contigs were binned with MetaBAT v2.12.1 (Kang et al., 2019) with a minimum contig length of 1500 nt. Binning quality was assessed using the CheckM v1.2.3 lineage work flow and quality commands (Parks et al., 2015). Bins of high (> 90% completeness, < 5% contamination) or medium (>50% completeness, < 5% contamination) quality were

retained for further processing (Pasolli et al., 2019). Bins were clustered based on average nucleotide identity (Shaw & Yu, 2023) and mutation distance (Ondov et al., 2016) by dRep v3.5.0 (Olm et al., 2017) to reduce redundancy. Representative genomes of each cluster were assessed for taxonomic placement using the GTDB-TK v2.4.0 (Chaumeil et al., 2022) classify work flow with the r220 representative database release (Parks et al., 2022). Genomes not classified down to the species level were assigned arbitrary values for the sake of identification and comparison. The representative MAGs were concatenated and built into a Bowtie2 database. The original AR reads from all American and comparative populations were aligned with –very-sensitive and -N 1 flags, duplicate alignments and those with a mapping quality < 37 were removed and filtered bam files were sorted and indexed with SAMtools. The abundance of each MAG was assessed with a custom script. The merged abundance table was visualized in R using the ggplot2 (v3.5.1) (Wickham, 2016), tidyverse (v2.0.0, https://r4ds.hadley.nz/), and compositions (v2.0-8, http://www.stat.boogaart.de/compositions/) packages.

3.3.9 Phylogenetic Reconstruction

MAGs generated from ancient American oral metagenomes were contextualized within an evolutionary framework to assess hidden species diversity. Phylogenies were constructed using both alignments of core genes assessed from a pangenome analysis and alignments of draft genomes obtained from aligning short reads to representative MAGs.

Pangenome Analysis and Core Gene Alignment

Three MAGs that either had high prevalence across populations or were clinically or historically relevant were queried for pangenome analysis. This included *Streptococcus sinensis*, Anaerolineaceae bacterium oral taxon 439 (GTDB: Flexilinea sp001717545), and *Eubacterium minutum* (GTDB: *Hornefia minuta*). Genes were predicted with prokka v1.14.6 (Seemann, 2014) for the representative MAGs and each genome within its secondary cluster (> 95% ANI) after dereplication. Published genomes downloaded from NCBI were also included for pangenome analysis. A pangenome was predicted using roary v1.007002 (Page et al., 2015) on all genome feature format (gff) files supplied by prokka. The parameters -e and -n were called to generate a multi-sequence alignment (MSA) file using MAFFT (Katoh et al., 2002). A maximum likelihood tree of the core gene alignment, prepared from genes shared across 99% of genomes, was built with IQTree with the following parameters: -T AUTO -m MFP -B 1000 -alrt 1000 -bnni. Phylogenetic trees with bootstrap support were visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

MAG Reference-based Alignment and Variant Calling

To capture hidden diversity of ancient American populations, original AR reads from the samples included in the *de novo* assembly pipeline, as well as comparative datasets described in *Comparative populations* and published genomes from NCBI, were aligned to the three MAGs derived from these ancient populations using BWA aln in the same manner described in *3.3.5 Source Tracking and Ancient Authentication* (Li & Durbin, 2009). Mapped SAM files were converted to BAM, sorted, indexed, and deduplicated with SAMtools

(Danecek et al., 2021) and genome statistics were computed with Qualimap v2.2.1 (Okonechnikov et al., 2016). aDNA damage was quantified and the deduplicated alignment file was rescaled with MapDamage v2.2.1 (Jónsson et al., 2013). Two nucleotides were trimmed at either end of the reads using BamUtil v1.0.15 (Jun et al., 2015) to prevent DNA damage from interfering with variant calling. A VCF file was then produced from the clipped bam file using VarScan v2.3.9 (Koboldt et al., 2009) mpileup2cns with the following parameters: --min-coverage 5 –min-reads2 3 –min-avg-qual 37 –min-var-freq 0.2 –min-freq-for-hom 0.9 –p-value 1 –strand-filter 0.

The resulting strains were included in the phylogeny if the alignment resulted in at least 30% of the MAG reference covered, a depth of five times, and a heterozygosity ratio < 1. VCF files were merged with BCFtools v1.17 (Danecek et al., 2021) and only bi-allelic SNPs were retained. Genomic regions annotated by prokka to be phage-related, mobile, or have variable copy numbers were excluded from the merged VCF file and insertion-deletion and SNP positions were removed if they had < 90% coverage. This value was decreased to 80% for Anaerolineaceae bacterium oral taxon 439 due to poor bootstrap support (<95%). The VCF file was converted to a multi-FASTA file excluding heterozygous sites and a maximum likelihood tree was built using IQTree with the same parameters described in 2.9.1. The final tree with bootstrap support was visualized in FigTree.

3.4 Results

3.4.1 Oral metagenome reconstruction and authentication

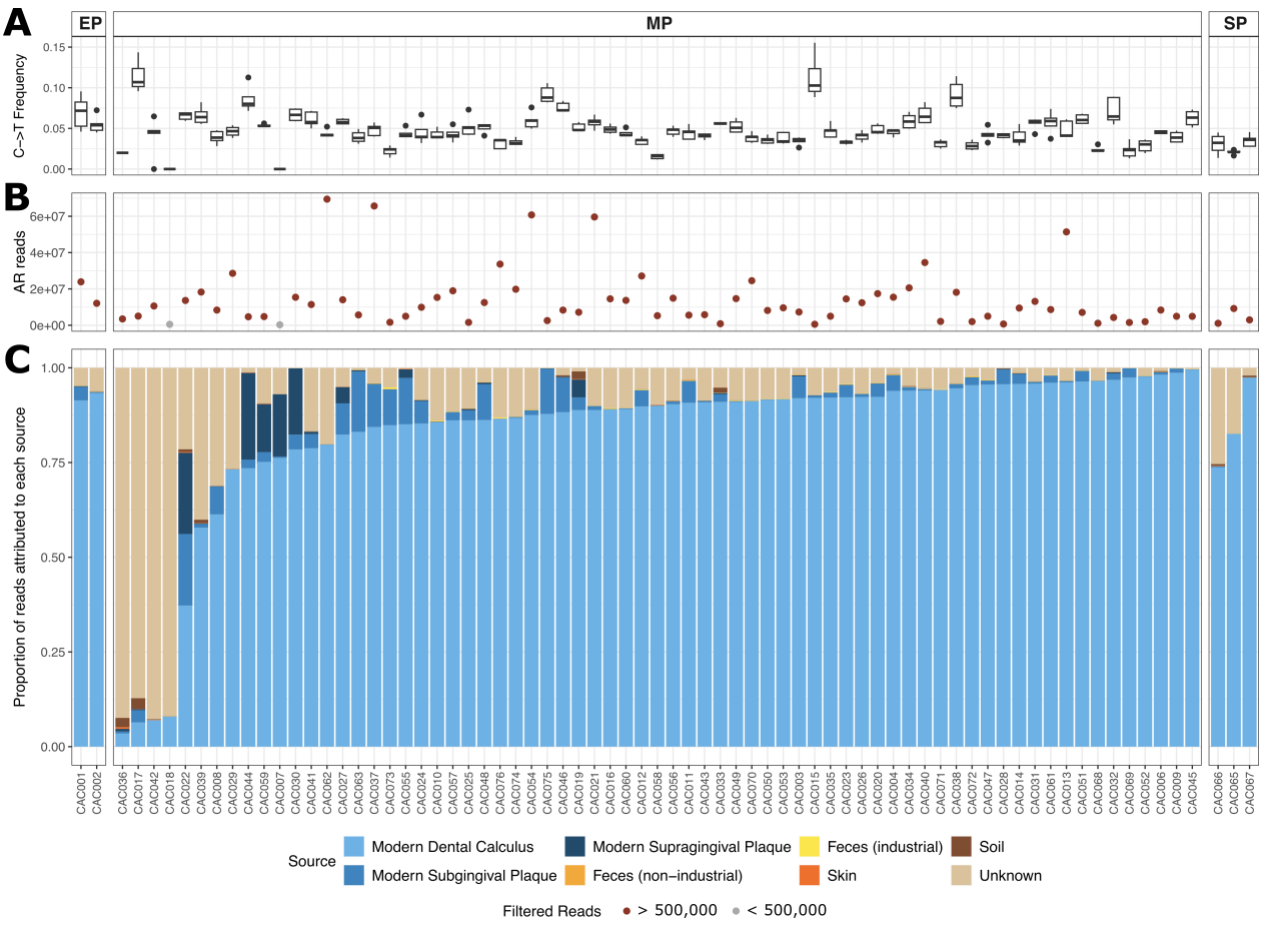


Figure 3.1 Summary plots.

A) Frequency of C to T transitions at first position of 5` end of DNA reads mapped to oral taxa (Anaerolineaceae bacterium oral taxon 439, an unclassified Firmicutes species (SGB71327), *Desulfobulbus oralis*, *Actinomyces dentalis*, and *Methanobrevibacter oralis*). Detailed line graph of 5` end damage frequencies included in Supplementary Figure 5. B) Number of AR reads for each sample in the SourceTracker2 plot. Grey circles indicate < 500,000 reads. C) Bar plot depicting proportions of source environments in each sample as determined by SourceTracker2 (Knights et al., 2011). (EP = Early Period, MP = Middle Period, SP = Mission Period)

Dental calculus was recovered from 75 Ancestors spanning 13 archaeological sites in the San Francisco Bay Area over a period of ~3,400 years. Nine calculus samples were

processed previously for shotgun metagenomic sequencing (Li, 2020), and 67 new samples (including an abscess scrape) were processed in this study. Overall, shotgun metagenomes were recovered from 76 samples, with sequencing depth ranging from 403,490 to 94,109,461 reads. Between 2% and 96% of reads were retained after trimming and merging with AdapterRemoval and a duplication rate of between 1% and 95.7% was observed. Deduplicated reads were mapped to the human reference genome (T2T-CHM13v2.0) to quantify host endogenous DNA content and detect nucleotide misincorporations indicative of ancient DNA damage. From those samples with > 1000 reads mapping to the human genome, cytosine to thymine (C to T) transitions were observed at an average rate of 5.3% at the terminal 5` nucleotide. Only reads not mapping to the human reference genome were retained for downstream analysis and are hereafter referred to as "analysis ready" or "AR" reads (Figure 3.1B).

Source tracking of taxonomic composition was employed to determine the proportion of AR reads possibly derived from contamination. The rate of classification for skin and soil sources was greater through the exact *k*-mer matching classification algorithm rather than the gene-marker based method (Chapter 2 – Figure 2.2), so taxonomic abundance profiles generated with Kraken2 (Wood et al., 2019) were used for source tracking rather than MetaPhlAn4 (Blanco-Míguez et al., 2023). Of the 76 metagenomes queried, 72 had greater than 50% of the metagenome attributed to an oral source, either from modern dental calculus, or supra- or subgingival plaque, and not from other potential contaminants, such as fecal material, soil, or skin (Figure 3.1C). The four

metagenomes with < 25% attributed to an oral source environment were removed from subsequent analyses.

AR reads were aligned to the five most abundant oral microbial genomes, including Anaerolineaceae bacterium oral taxon 439, an unclassified Firmicutes species (SGB71327), *Desulfobulbus oralis*, *Actinomyces dentalis*, and *Methanobrevibacter oralis*, to detect DNA damage diagnostic of ancient samples. The rate of C to T transitions at the 5` end was quantified using mapDamage2 (Jónsson et al., 2013) and the frequencies from the first position of all alignments were plotted for each sample (Figure 3.1A). Each metagenome with oral microbial preservation and greater than 500,000 AR reads had an ancient signature, defined as such if samples with greater than 600 uniquely aligned reads to the reference had at least a 1% frequency of C to T transitions at the first position. This resulted in 69 reconstructed ancient oral metagenomes, making this the largest oral microbiome dataset from a single geographic area in the Americas to date, ancient or otherwise.

3.4.2 Taxonomic classification

Taxonomy was assigned to AR reads through a dual approach, using the marker-gene based short read aligner, MetaPhlAn4 (Blanco-Míguez et al., 2023), for compositional data analysis, and the exact *k*-mer matching program, Kraken2 (Wood et al., 2019), for source tracking and eukaryotic DNA screening. This process is described previously (2.3.6 Taxonomic profiling), but the marker-gene based approach of MetaPhlAn4 is algorithmically similar to the functional profiling tool also used in this analysis, HUMAnN3

(Beghini et al., 2021). Abundances are automatically normalized by genome length and the marker-gene approach reduces false positive rate (Blanco-Míguez et al., 2023; Velsko et al., 2018).

Taxonomic profiles were filtered of rare taxa, keeping those with a relative abundance of at least 0.005% in at least 20% of samples. To avoid the influence of potential environmental or laboratory contaminants in microbial community interpretations, profiles were limited to oral-associated microorganisms classified as such through the Human Oral Microbiome Database (HOMD) or identified in modern oral environments (Escapa et al., 2018). As metagenomic data are compositional in nature, relative abundances were CLR transformed prior to further analysis. The most abundant oral taxa identified across the dataset on average include *Actinomyces dentalis*, Anaerolineaceae bacterium oral taxon 439, *Desulfobulbus oralis*, and unclassified Synergistaceae and Firmicutes MAGs previously isolated from modern and historic oral metagenomes, respectively (Blanco-Míguez et al., 2023; Pasolli et al., 2019).

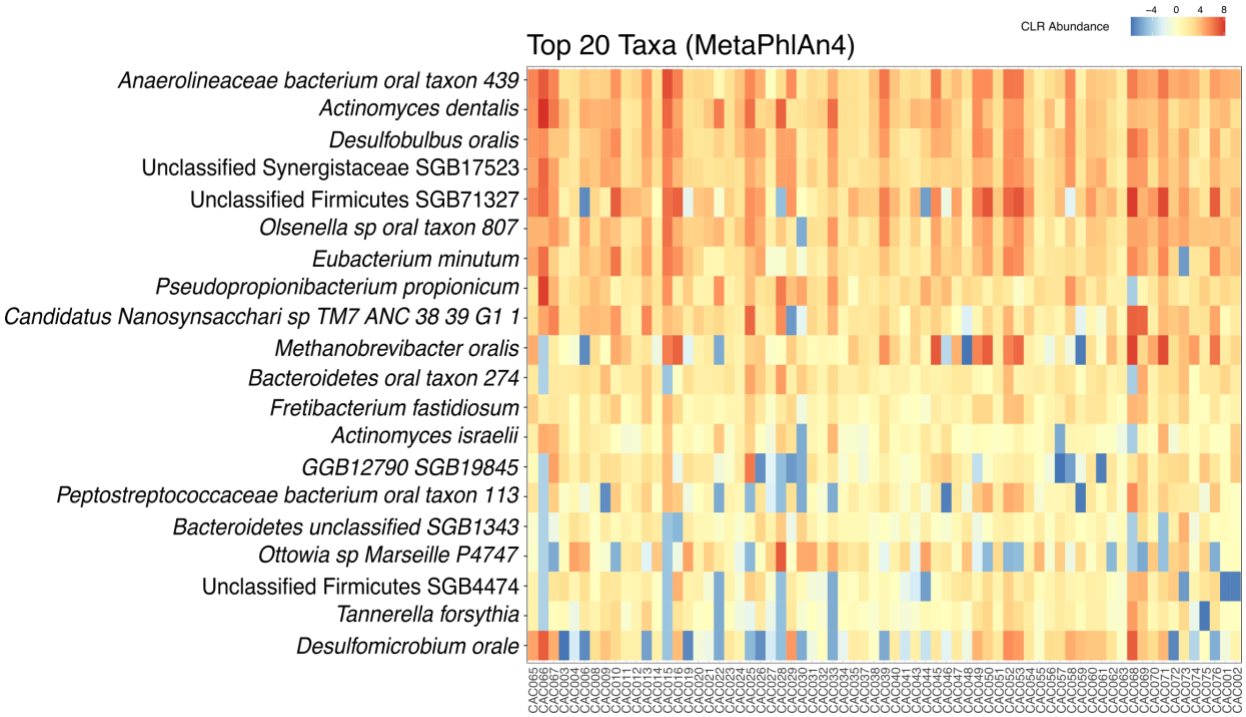


Figure 3.2 Taxonomic abundance heatmap.

Top 20 most abundant species identified in CLR-transformed microbial profiles classified by MetaPhlAn4. Heatmap generated using ggplot2 (Wickham, 2016).

To reduce dimensionality of the community composition and observe any impact of metadata on sample clustering, the CLR-transformed table was ordinated into a principal component (PCA) plot and a permutated multivariate analysis (PERMANOVA) was performed. Extraction and library batch were significantly correlated with PC ordination (p-value < 0.05), but archaeological site, total calculus, Ct value, sex, age, and disease status had no influence on sample clustering. However, when controlling for sequencing facility (OMRF vs Admera), neither extraction nor library batch showed significant correlation. Metagenomes were recovered from three SP individuals and from two from individuals who predate the practice of intensive acorn consumption (EP). The community composition of these groups fell within the PCA distribution of the MP samples and was not significantly correlated with PC ordination. Overall, this analysis suggests that taxonomic composition

of these samples is not influenced by diet, oral disease, temporal period, or archaeological site, but batch effects due to sequencing facilities likely play a role driving observed variation.

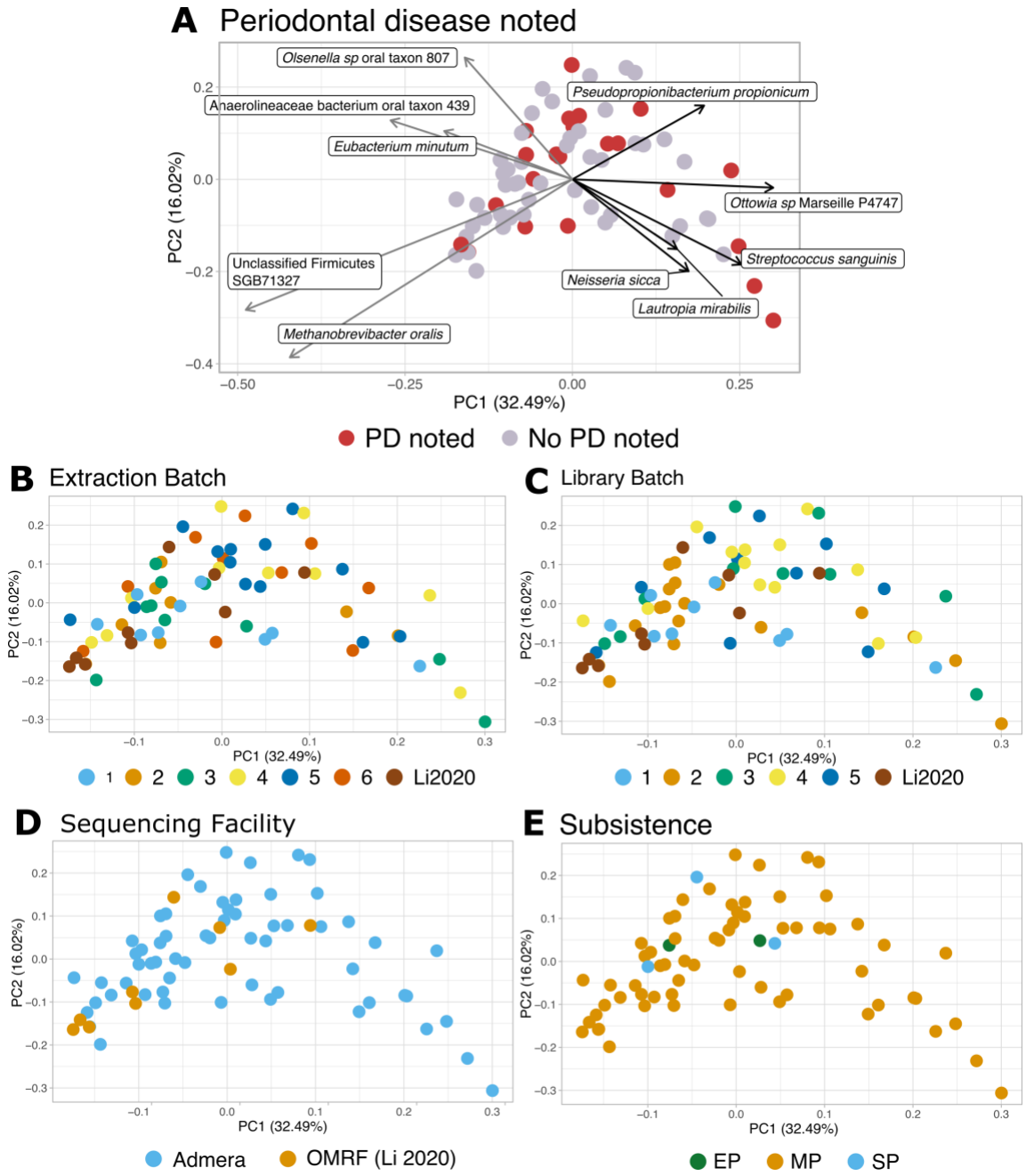


Figure 3.3 Principal component analysis of community composition.

A) PCA colored by disease status. Biplot indicates top taxa driving separation in PCA space. B-C) Extraction and library batches. Li 2020 indicates samples run previously by co-

author JL. D) Colored by sequencing facility. E) Early, Middle, and Mission era samples noted.

This taxonomic profile was compared to other pre-European contact American populations, including four previously published profiles of Muwekma Ohlone Ancestors (Eerkens et al., 2018), Ancestral Maya in Belize from before (PM-MRSC) and after (SM-MRSC) the adoption of maize as a dietary staple (Chapter 2) and agricultural Wichita Ancestors from Oklahoma (Honap et al., 2023), as well as an historic population from the UK (Velsko et al., 2019) and modern individuals from Spain (Fellows Yates (B) et al., 2021; Velsko et al., 2019). These comparative metagenomes were combined, and the merged abundance table was CLR-transformed and ordinated in a principal coordinate plot to observe patterns of sample clustering (Figure 3.4). Metagenomes from both the PM-MRSC (the oldest samples), and the modern individuals form tight clusters with no overlap in ordinal space. The historic individuals from the Radcliffe Infirmary Burial Ground are more widespread but overlap more with the modern European samples than with the American populations. All American populations overlap, including the PM-MRSC samples, which cluster tightly. Each population represented in the comparative analysis has distinct characteristics that either contrast or overlap with qualities of California Ancestors, highlighting aspects of the California metagenomes that can be attributed to their unique subsistence style.

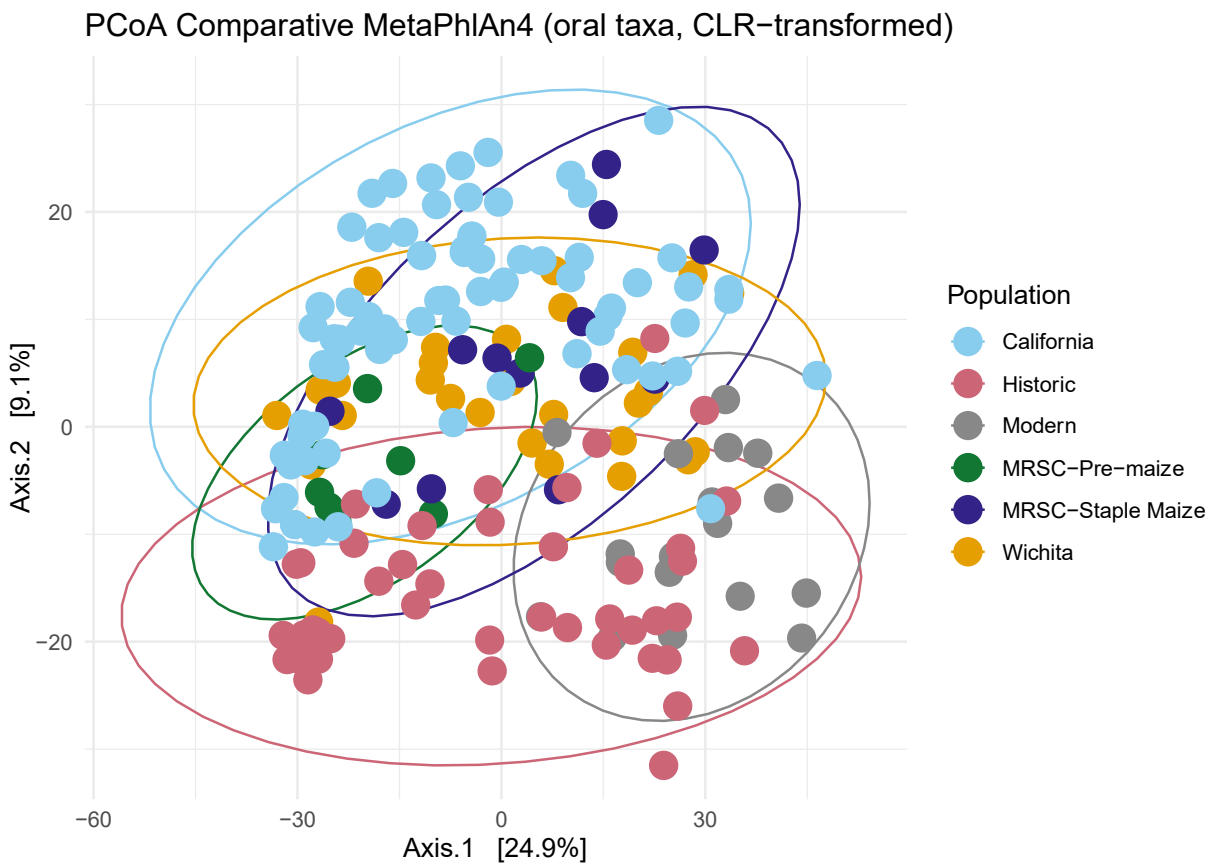


Figure 3.4 PCoA of taxonomic profiles between comparative populations.

Ellipses encircle multivariate t distribution. Euclidean distance matrix computed with phyloseq (McMurdie & Holmes, 2013) and ordinated in ggplot2 (Wickham, 2016).

The top 20 most abundant taxa across samples in each population were plotted together in a heatmap (Figure 3.5). Abundant taxa shared across populations are *Actinomyces* species and its symbiont, Candidatus Nanosynsacchari sp. TM7, Anaerolineaceae bacterium oral taxon 439, *Pseudopropionibacterium propionicum*, *Olsenella* sp. oral taxon 807, and *Desulfobulbus oralis*. *Capnocytophaga, Rothia,* and *Neisseria* species populate the modern oral microbiomes but are mostly absent in the PM-MRSC and sparse in the California, Wichita, and SM-MRSC. Some taxa absent in the modern and historic European populations include *Propionibacterium* sp. oral taxon 192 and *Actinomyces oricola*. Notably, three unclassified Firmicutes species level genome

bins (SGBs) recovered from historic individuals (SGB70729, SGB71326, and SGB71327) (Pasolli et al., 2019; Velsko et al., 2019), are primarily identified in pre-modern populations but are absent in the modern individuals. The modern individuals also lack *Methanobrevibacter oralis* and, along with the PM-MRSC, have reduced prevalence of *Streptococcus sinensis*. Random forest analysis indicates that these differentially abundant taxa across groups are predictive of group divisions (Figure 3.6).

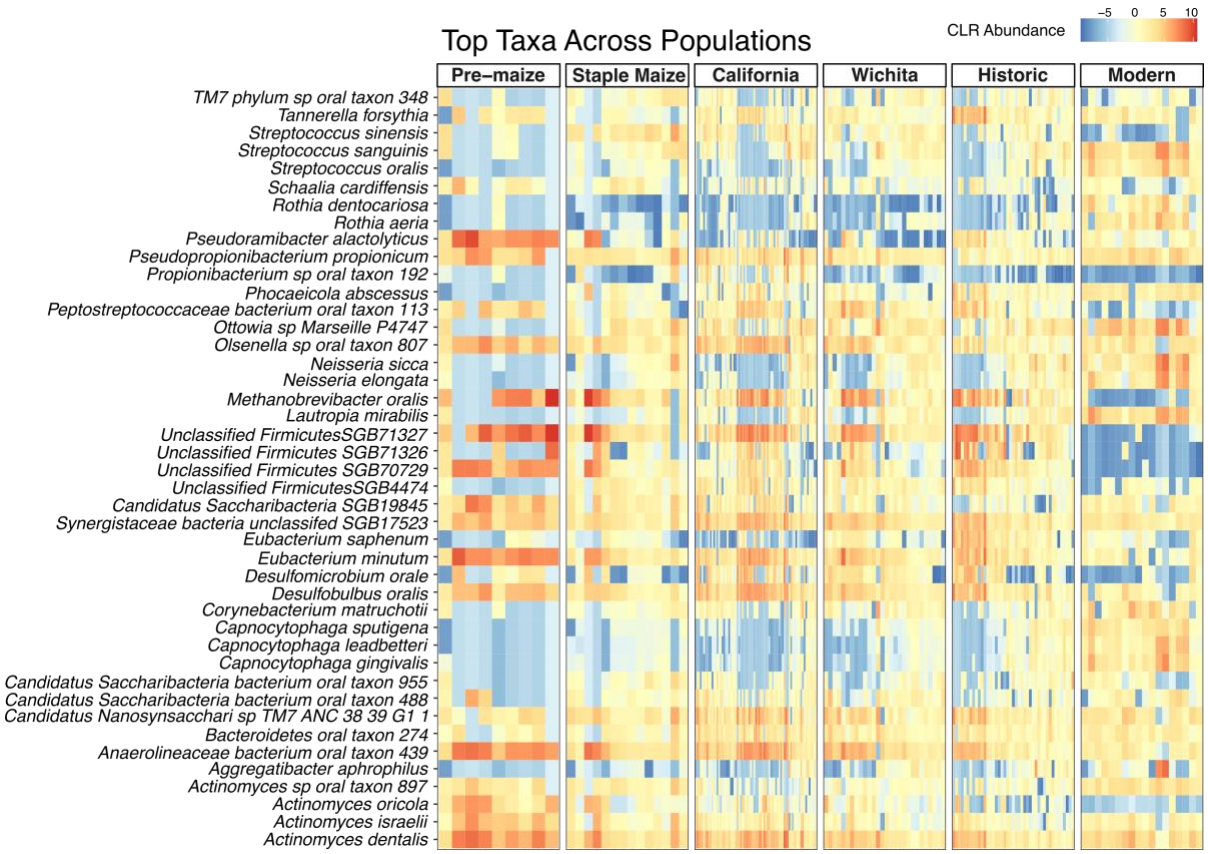


Figure 3.5 Heatmap of abundant microbes across populations.

A total of 43 taxa are shown here which include the 20 most abundant CLR-transformed taxa from each population determined using MetaPhlAn4.

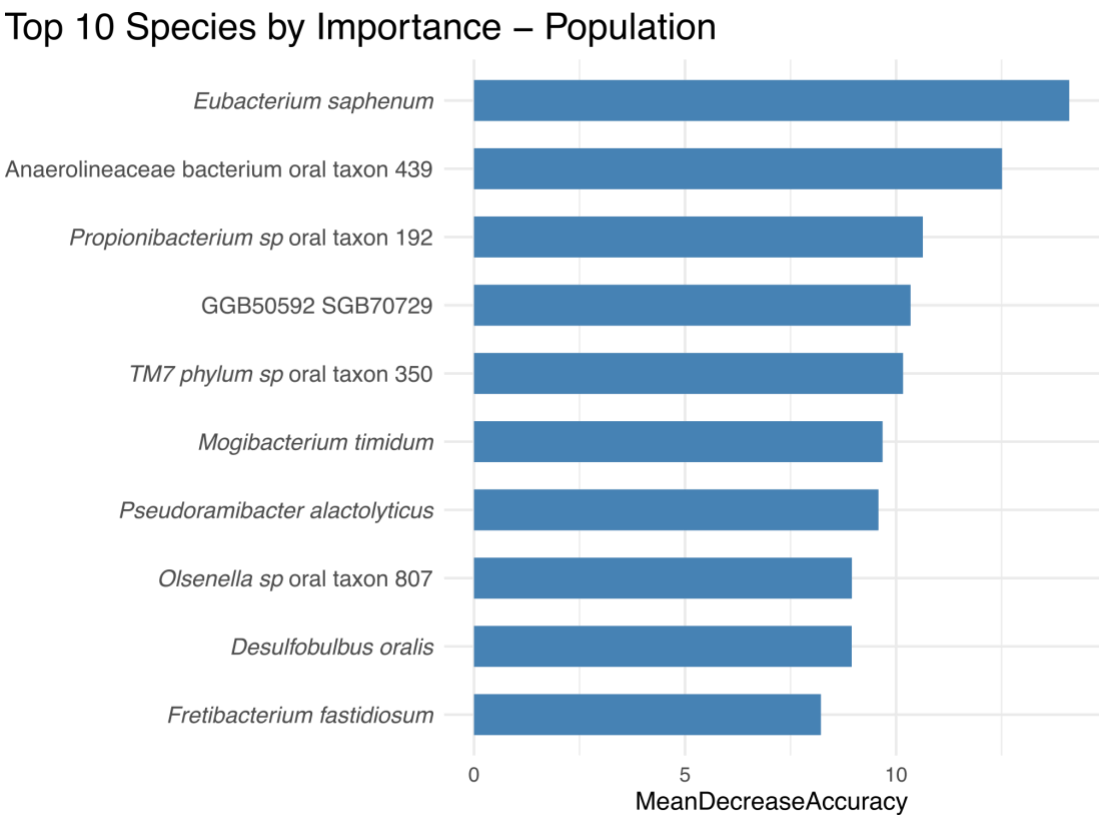


Figure 3.6 Top species with high variable imporance.

Top ten species determined by random forest classifier to be predictive of population separation. Species ordered by mean decrease accuracy, or likelihood of decreasing model accuracy if excluded from the classification.

3.4.3 Functional Profiling

To investigate the effects of dietary habits on microbial dynamics, Ancient California and comparative metagenomes were characterized for microbial functional potential. The comparative populations described above were functionally annotated with HUMAnN3 (Beghini et al., 2021) and organized into metabolic categories through KEGG orthologous (KO) functional groups (Kanehisa et al., 2014; Kanehisa & Goto, 2000). KO functions stratified by contributing species were CLR-transformed to accommodate the compositional nature of the data (Gloor et al., 2017; Quinn et al., 2019). Euclidean

distances were then computed for a principal coordinate (PCoA) plot (Figure 3.7B). Each population has distinctive characteristics, and thus structuring is observed in this ordination. The samples from Ancestral Maya predating maize-based agriculture (PM-MRSC) are the oldest (> 4,000 YBP) (Chapter 2) and cluster tightly in the center of the plot, overlapping with the other pre-modern populations. The samples from Ancestral Maya with maize as a dietary staple (SM-MRSC) cluster with samples from the California Ancestors, including those published previously (Eerkens et al., 2018), and both of these share ordination space with the samples from the Wichita Ancestors. The modern calculus samples (Fellows Yates (B) et al., 2021; Velsko et al., 2019) distinctly overlap with the those from agricultural populations (historic Radcliffe and Wichita Ancestors). The five California samples not interred during the MP are noted on both PCoA plots in Figure 3.7. Notably, the SP samples in dark maroon have a more positive ordination along Axis 2 with respect to the EP samples in the functional plot (Figure 3.7B). The EP samples in light green reside at the fringes of the Ancient California, SM-MRSC, and Wichita populations. The SP samples cluster closer to the central PM-MRSC and agricultural Wichita and Radcliffe samples. This pattern is not observed with such distinction in the taxonomic PCoA (Figure 3.7A).

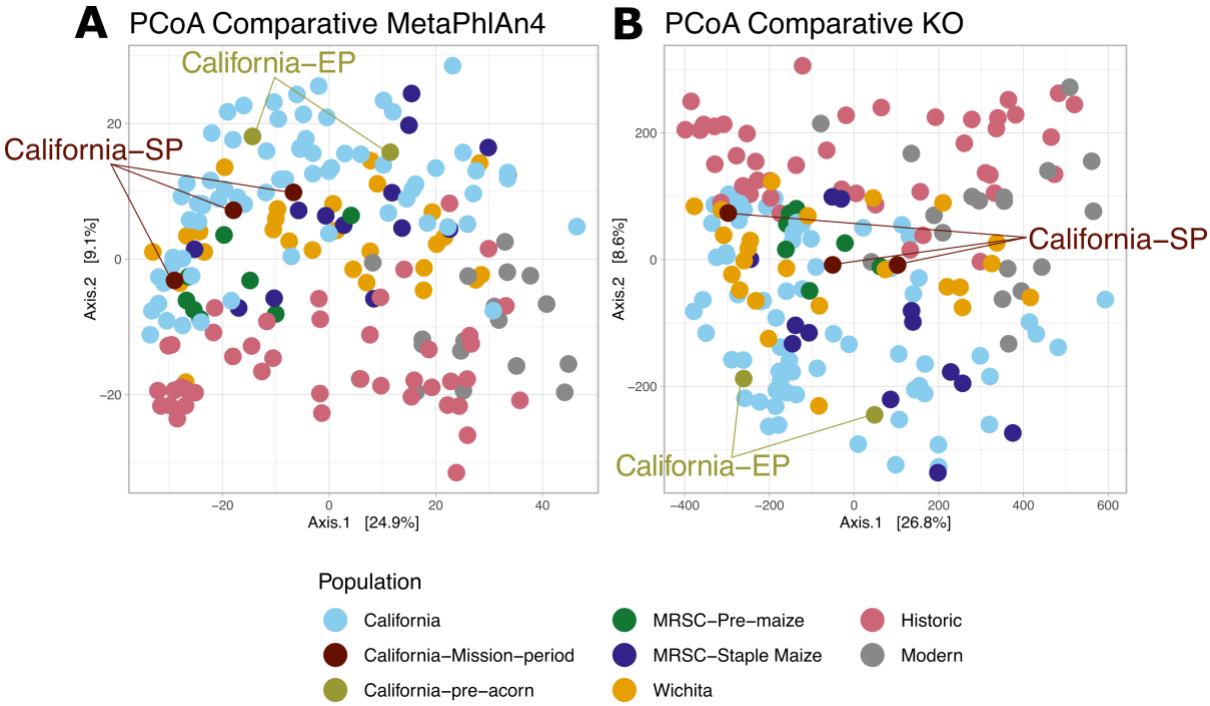


Figure 3.7 PCoA between comparative populations.

A) PCoA of taxonomic composition profiled with MetaPhlAn4. B) PCoA of Gene family tables from HUMAnN3 were normalized to copies per million and regrouped into KO functions. Both figures depict Euclidean distances of CLR-transformed data ordinated as a principal coordinate plot with phyloseq (McMurdie & Holmes, 2013) and ggplot2 (Wickham, 2016).

Given the distinct separation across ordinate space for functional properties, we investigated specific metabolic categories that may be contributing to such a pattern. Here, given the unique consumption of high-starch food items despite their non-agricultural lifestyle, we intuitively focused on functions related to starch and sucrose metabolism. The stratified functional table was subset to include only KO functions categorized within the starch and sucrose metabolism pathway (Kanehisa et al., 2014). These tables aggregated as either function or contributing species were then modeled with a random forest algorithm to determine the most differential features of either table. Figure 3.8 displays the relative abundances of the top ten most important functions contributing

to the uniqueness of the comparative populations (mean decrease gini) determined by the random forest classifier.

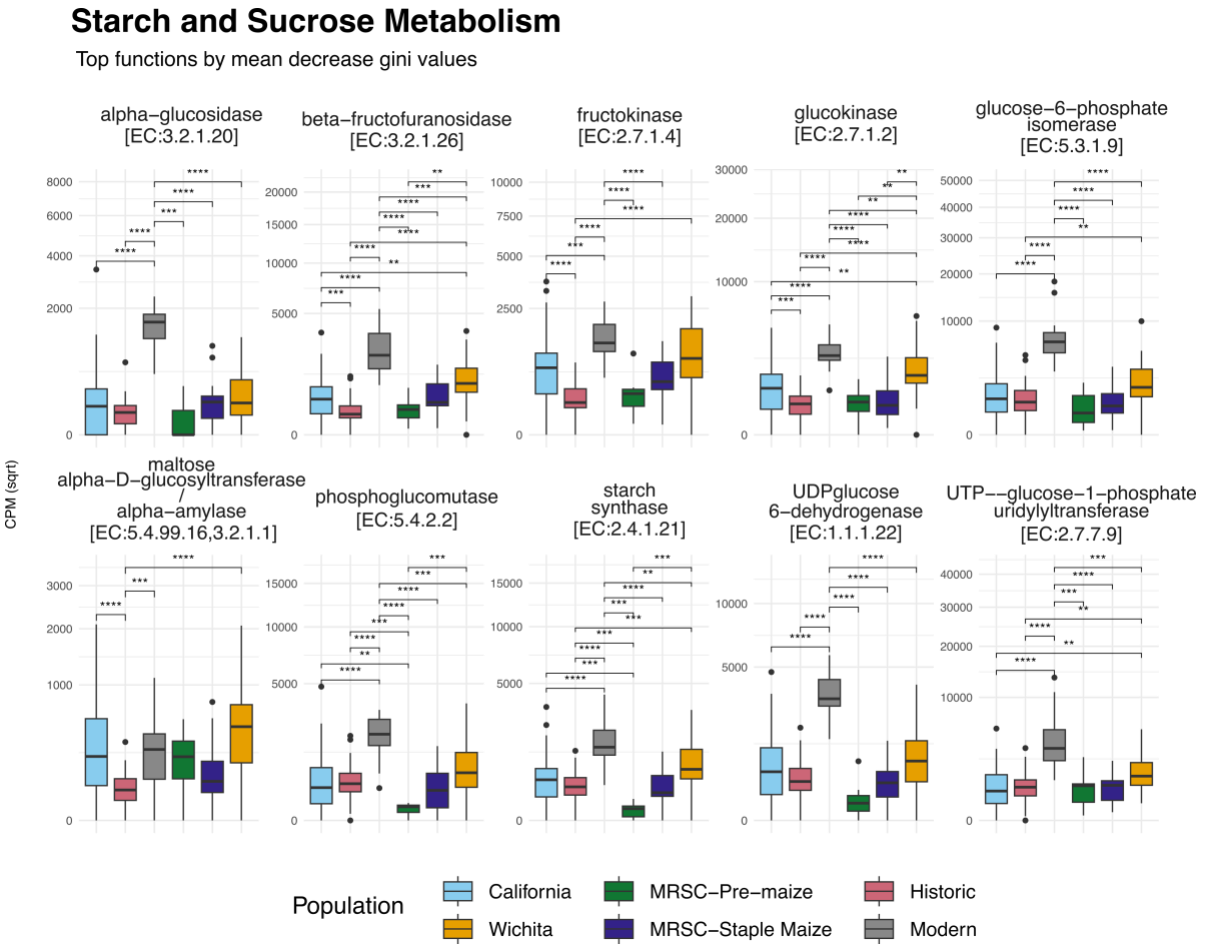


Figure 3.8 Boxplots of starch and sucrose metabolic functions.

Stratified functional tables were subset to KO functions classified as "starch and sucrose metabolism". Functions were determined by a random forest model to have high mean decrease gini values in differentiating populations. The y-axis is scaled by square-root transformation for aesthetic purposes. Pairwise-Wilcoxon test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).

Fructokinase has the greatest variable importance in distinguishing between the comparative populations with respect to starch and sucrose metabolism. The relative abundance of taxa encoding fructokinase is significantly greater in modern dental calculus compared to all other populations, except the Wichita Ancestors. *Streptococcus sanguinis*

is the top contributor to this function in modern populations, as opposed to Anaerolineaceae bacterium oral taxon 439 in pre-modern populations (Supplementary Figure 6). Multiple *Streptococcus* species contributed to this function in pre-modern populations, including *S. sanguinis*, *S. sinensis*, *S. salivarius*, *S. oralis*, *S. gordonii*, and *S. cristatus*. The abundance of fructokinase from *S. oralis* is negligible in the California Ancestors, while the historic Radcliffe and Wichita Ancestors lack *S. salivarius*. The PM-MRSC do not have *Streptococcus* species contributing to this function.

Actinomyces sp. oral taxon 897 has high variable importance for starch and sucrose metabolism between populations but does not contribute to this function in either hunter-gatherer population (Muwekma Ohlone or the PM-MRSC) (Supplementary Figure 7). This taxon contributes to five of the 10 functions of variable importance, including UDPglucose 6-dehydrogenase, glucokinase, alpha-glucosidase, beta-fructofuranosidase, and glucose-6-phosphate isomerase. In lieu of this taxon, American pre-agricultural starch and sucrose metabolic functions are sourced from Anaerolineaceae bacterium oral taxon 439, *Desulfomicrobium orale*, *Desulfobulbus oralis*, *Tannerella forsythia*, and *Pseudopropionibacterium propionicum* (sometimes classified as *Arachnia propionica* (Scholz & Kilian, 2016)). The relative abundance of *P. propionicum* in California metagenomes is similar to all other American agricultural groups, but the abundance of *D. oralis* is more similar to the Wichita than the SM-MRSC.

Despite the distinct characteristics differentiating the comparative populations, the greatest difference observed is between each pre-modern population and the modern metagenomes. For almost all starch and sucrose metabolic functions, the modern

samples have significantly greater abundance than any other population. With few exceptions, the California starch and sucrose metabolic functions are surprisingly not significantly different from those of the historic Radcliffe metagenomes, despite the difference in geographic location, time, and subsistence strategy. However, when looking at species contributing to these functions (Supplementary Figure 7), the relative abundances of all but two are significantly different between these two populations. This suggests a resilience in starch and sucrose metabolic function despite differential taxonomic abundance.

3.4.4 Metagenome Assembly and Binning

Three ancient oral metagenomic datasets from the Americas were queried for the possible reconstruction of ancient metagenomically assembled genomes (MAGs). Metagenomes from all three populations, the California, Ancestral Maya, and Wichita, were individually assembled, binned, and assessed for quality. Of the 119 individual metagenomes assembled, 117 resulted in at least one contig greater than 1000 nucleotides (nt) and 21 samples elicited at least a medium-quality bin, > 50% completeness and < 5% contamination reported by CheckM (Parks et al., 2015; Pasolli et al., 2019). There was a total of 372 high-quality bins (>= 90% completeness, < 5% contamination) and 490 medium-quality bins (50% - 89% completeness, < 5% contamination). High and medium-quality bins were used for subsequent analyses unless otherwise indicated. Bins were de-replicated with dRep (Olm et al., 2017) and an average nucleotide identity (ANI) threshold of 95% resulted in 133 representative genomes.

Classification by GTDBTk (Chaumeil et al., 2022) resulted in 44 bins classified to the species level, whereas 89 bins could not be classified at the species level and 16 remained unclassified even at the genus level.

There is a positive correlation between the quality of metagenome assembly and the number of high- to medium-quality bins produced. Assembly metrics such as L90 (the number of contigs that comprise 90% of the genome/assembly), maximum contig length, and total number of contigs are all strongly correlated (> 0.5 Pearson correlation) to the number of retained bins. N90 (length of the contig above which comprises 90% of the genome/assembly), standard deviation of GC content, and N50 (length of the contig above which comprises 50% of the genome/assembly) are all moderately correlated (0.3-0.5), and average GC content (ratio of guanine and cytosine nucleotides) and L50 (the number of contigs that comprise 50% of the genome/assembly) are weakly correlated (0-0.3).

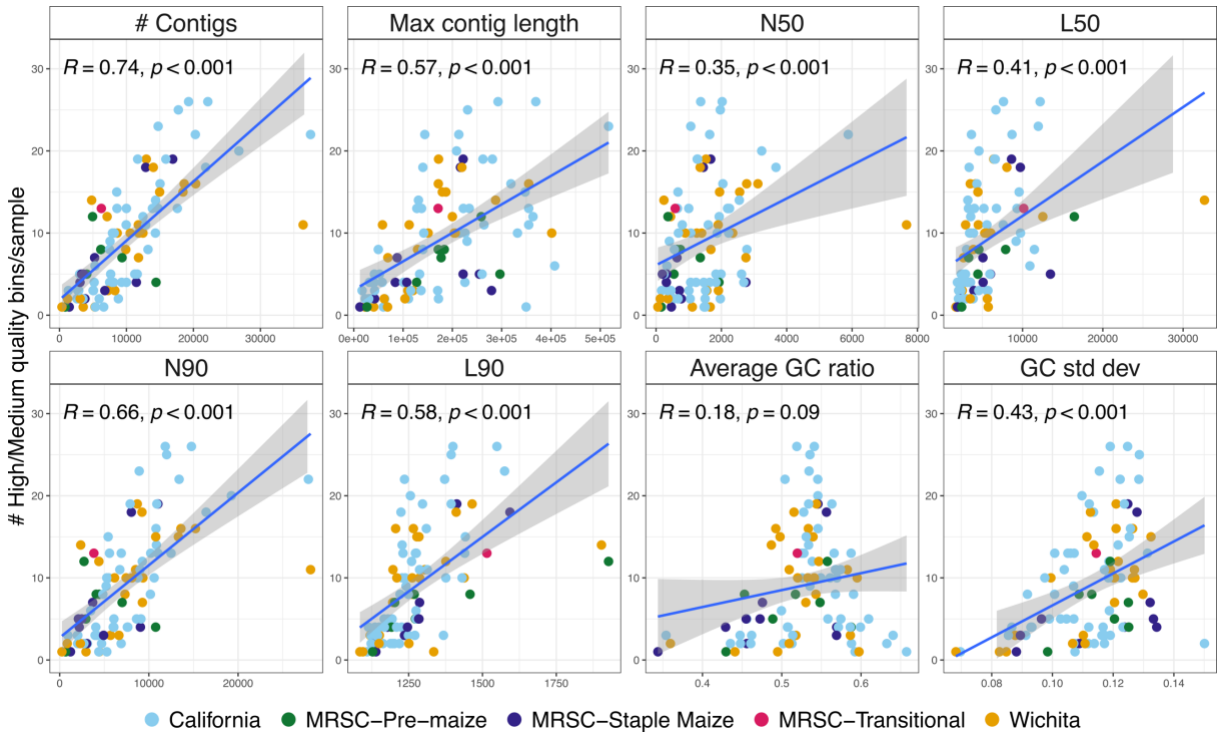


Figure 3.9 Influence of assembly quality on MAG construction.
Linear models for all individual assemblies. Depicts correlation between the measure of assembly quality and the number of high-/medium-quality bins produced.

The medium- to high-quality bins were taxonomically classified using the GTDBTk classify_wf pipeline with the r220 database release. To reduce redundancy, bins were dereplicated, resulting in 133 different clusters at a threshold of 95% ANI. The number of bins per cluster ranged from 1 to 60, with an average of 7. The most populated clusters were classified as *Methanobrevibacter oralis*, Anaerolineaceae bacterium oral taxon 439, and an unclassified Anaerovoracaceae (Eubacteriales). The genome in each cluster with the highest score, calculated internally based on N50 and genome length, was recognized as the representative genome of that cluster. The representative genomes were then built into a Bowtie2 database to measure the relative abundances of each genome in the original short-read metagenomes.

3.4.5 Genomic Reconstruction and Phylogeny

We recovered two MAGs from the assemblies that were either identified to be informative of population structure via random forest, or of high relative abundance across populations. This includes *Eubacterium minutum* (GTDB: *Hornefia minuta*), and Anaerolineaceae bacterium oral taxon 439 (GTDB: *Flexilinea sp001717545*). Additionally, through the dereplication pipeline, we identified one *Streptococcus sinensis* cluster that included 10 MAGs sharing at least 95% ANI. Published genomes included in each pangenome analysis had < 5% variation from other MAGs in the clusters. A pangenome

analysis was performed on the three reconstructed genomes to characterize and annotate core genes, or those shared across 99% of genomes in the cluster. Furthermore, AR reads from the comparative metagenomes described in 3.3.1 Study Population were aligned to the representative genome of each of these MAGs for a draft genome reconstruction through a reference-based approach followed by variant calling. Both the core gene and genome-wide SNP-based alignments were used for phylogenetic analysis.

		MAG reference alignment and variant calling		**Pangenome analysis and core gene alignment**	
MAG	**GTDB Classification (NCBI)**	**Number of genomes in alignment**	**Average heterozygosity ratio**	**Number of genomes in core alignment**	**Number of core genes**
CAC076.bin.39	*Hornefia minuta* (*Eubacterium minutum*)	59	0.15	37	384
WAMP26.bin.13	*Flexilinea sp001717545* (Anaerolineaceae bacterium oral taxon 439)	22	0.49	26	1286
MRSC-2072.bin.26	*Streptococcus sinensis*	2	0.275	14	95

Table 3.3 Statistics of MAGs included in pangenome analysis and reference-based phylogenies.

Reference-based alignments were limited to those where at least 30% of the reference was covered at least five times and the heterozygosity ratio was < 1, limiting the alignment to single strains. Each MAG had variable average heterozygosity ratios, with the lower rate corresponding to greater number of genomes included in the phylogeny (Table 3.3). Core gene alignments had improved bootstrap support if limited to high-quality MAGs with at least 90% completeness indicated by CheckM (Parks et al., 2015; Pasolli et al.,

2019). Although only medium-quality MAGs classified as *S. sinensis* were recovered from the ancient metagenomic datasets, the phylogeny had strong bootstrap support.

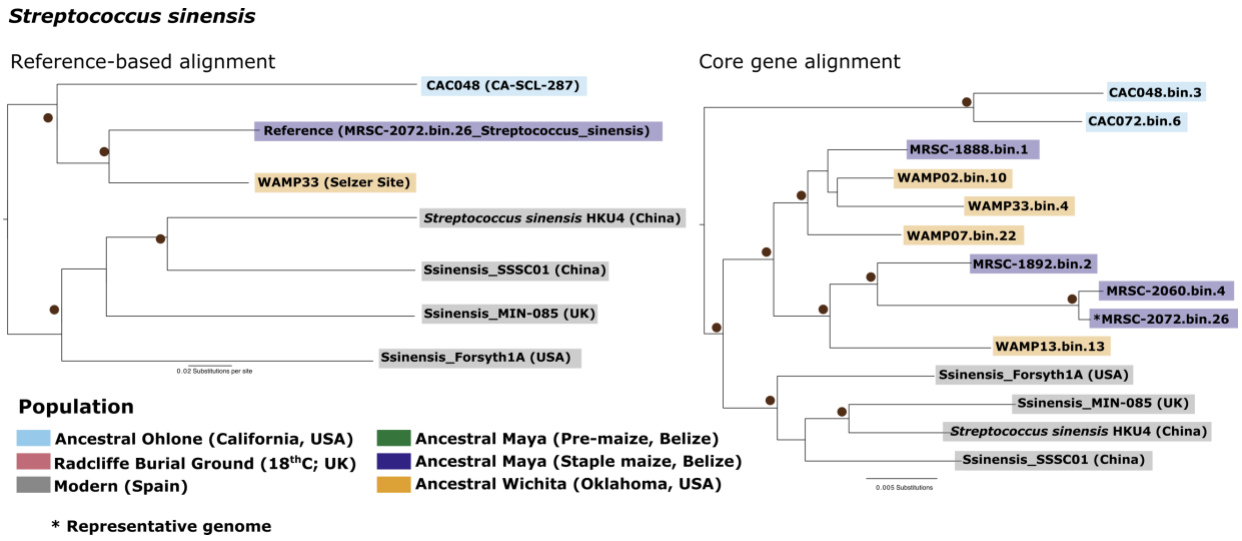


Figure 3.10 *Streptococcus sinensis* ML trees.

Reference-based alignment computed on 11,446 bi-allelic SNPs, allowing 10% missing data. Tree was built using the TVM+F+ASC model in IQ-Tree and midpoint rooted. Core gene tree built with 95 core genes (shared across 99% of samples) using the GTR+F+R2 model in IQ-Tree and midpoint rooted. Trees visualized using FigTree. Brown circles indicate > 95% bootstrap support using 1,000 ultrafast bootstrap replicates. Colored boxes indicate population MAG was isolated from. Isolate location of published genomes is indicated where applicable.

Streptococcus genomes have been difficult to reconstruct through reference-based methodologies (Chapter 1) (Velsko & Warinner, 2025). Here, we have generated a *S. sinensis* genome through metagenomic assembly and contig binning from ten different ancient metagenomes (Figure 3.10). The representative MAG from a SM-MRSC metagenome (MRSC-2072) was used as reference to align short-reads of comparative metagenomes, as well as modern published genomes. Out of 262 alignments to the reference MAG, 19 covered > 30% of the reference at least five times, 15 of which were metagenomes of agricultural individuals. The other four belonged to Muwekma Ohlone Ancestors. However, only two had a heterozygosity ratio < 1 indicating a high likelihood of

more than one strain in the aligned metagenomes. Alternatively, ten medium-quality MAGs were recovered through the assembly-based methodology. For both the reference-based and core gene multi-sequence alignment (MSA), the modern genomes cluster in a single lineage and the Muwekma Ohlone metagenomes are divergent from the SM-MRSC and Wichita genomes. The ANI between the Muwekma Ohlone MAGs and most others is slightly lower than the 95% for most other comparisons, though this is not consistent throughout the tree. For the core gene MSA, there is no geographic or temporal structuring of the SM-MRSC or Wichita genomes.

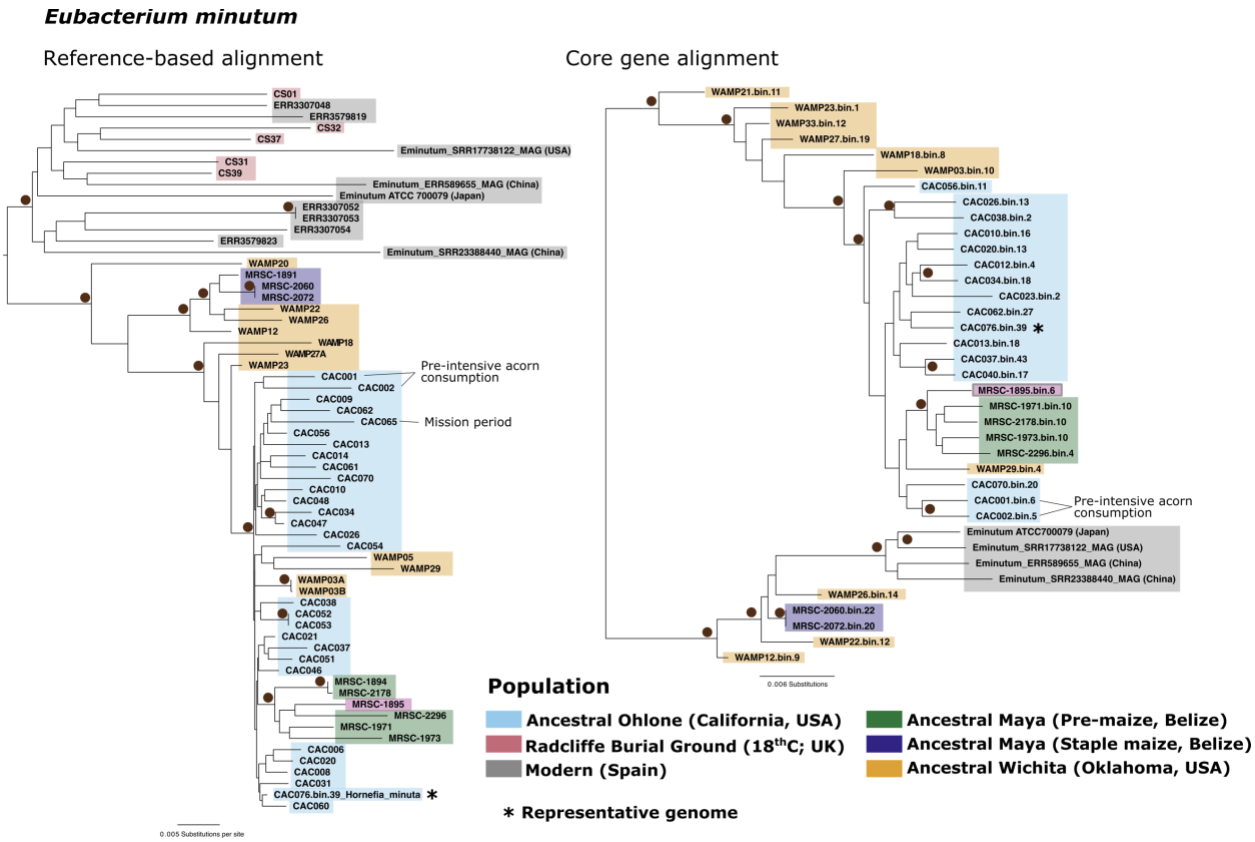


Figure 3.11 *Eubacterium minutum* ML trees.

Reference-based alignment computed on 2,935 bi-allelic SNPs, allowing 10% missing data. Tree was built using the TVM+F+ASC+R3 model in IQ-Tree and midpoint rooted. Core gene alignment limited to high-quality MAGs. ML tree built with 409 core genes using the GTR+F+R8 model in IQ-Tree and midpoint rooted. Trees visualized using FigTree. Brown circles indicate > 95% bootstrap support using 1,000 ultrafast bootstrap replicates.

Colored boxes indicate population from which MAG was isolated. Muwekma Ohlone metagenomes from before intensive acorn consumption and Amah Mutsun metagenomes from after European contact are indicated.

Eubacterium minutum is highly abundant across historic and ancient oral metagenomes (Figure 3.5) and is a significant driver of compositional structuring between metagenomes from before and after the adoption of maize as a dietary staple (Figure 2.4). Thirty-three high-quality genomes were used for a core gene MSA. The representative MAG recovered from an Muwekma Ohlone metagenome (CAC076) was used as a reference to reconstruct 59 draft genomes from comparative samples. The phylogeny constructed with the MAG as reference has two distinct lineages, with one containing all historic and modern strains from European metagenomes and all modern published genomes isolated from either Japan, China, or USA. The other linage contains all strains from pre-European contact American metagenomes in addition to one from the SP Amah Mutsun Ancestor (CAC065). Within the predominantly pre-contact American branch, Wichita strains are dispersed across the tree, irrespective of shared diet, burial site, or sample age. PM- and TM-MRSC strains cluster alongside all 28 California strains and are divergent from SM-MRSC strains. Two *E. minutum* strains were recovered from EP metagenomes. These two are basal to the remaining Muwekma Ohlone, PM- and Transitional-MRSC, and four Wichita strains.

The core genome phylogeny differs slightly in tree topology. There was no MAG recovered from the SP metagenome. The MAGs from those metagenomes predating acorn consumption (EP) are still basal to the other Muwekma Ohlone MAGs, but clusters with an individual from the MP. The PM-MRSC MAGs are on the same lineage as the Muwekma

Ohlone, but the SM-MRSC and three Wichita MAGs reside with the modern published genomes.

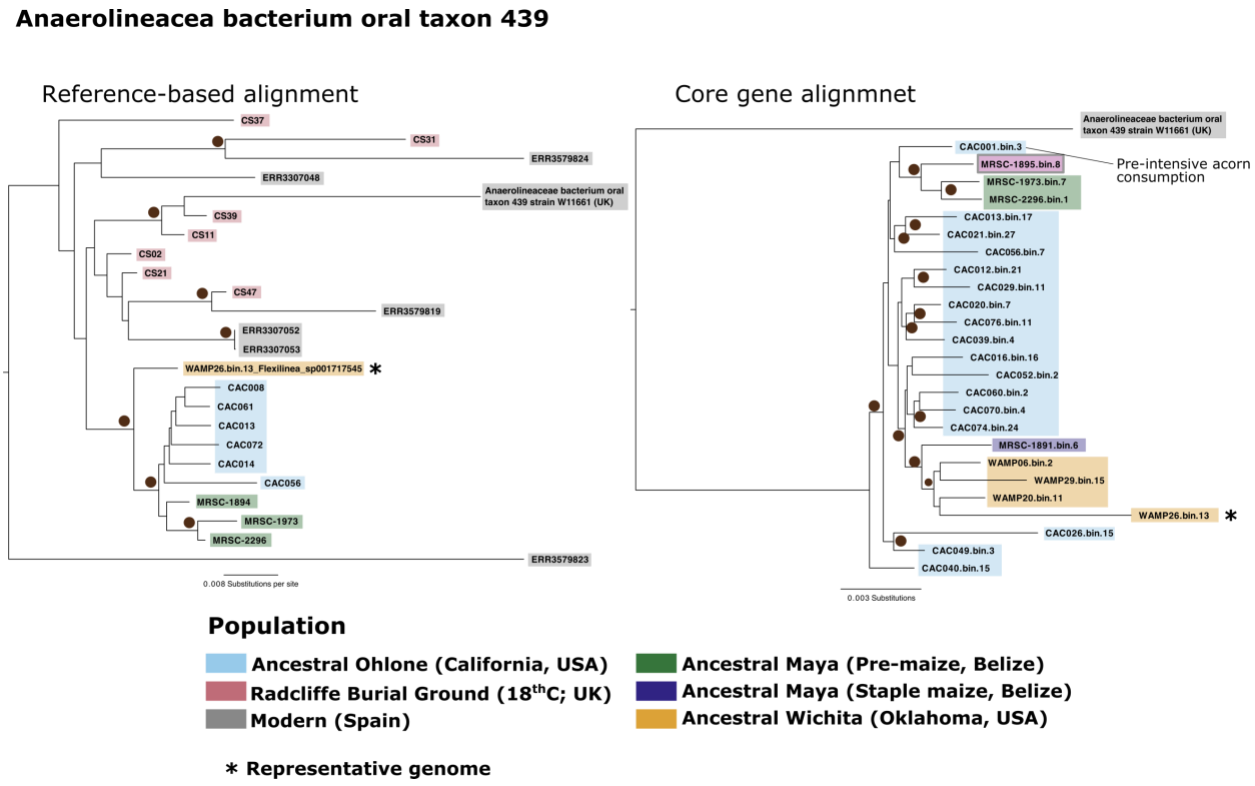


Figure 3.12 Anaerolineaceae bacterium oral taxon 439 ML trees.

Reference-based alignment computed with 3,607 bi-allelic SNPs, allowing 20% missing data. Tree was built using the TVM+F+ASC+R2 model in IQ-Tree and midpoint rooted. Core gene alignment was limited to high-quality MAGs. ML tree built with 1286 core genes using the GTR+F+R6 model in IQ-Tree and midpoint rooted. Trees visualized using FigTree. Brown circles indicate > 95% bootstrap support using 1,000 ultrafast bootstrap replicates. Colored boxes indicate population from which MAG was isolated. An Muwekma Ohlone metagenome from before intensive acorn consumption is indicated.

Anaerolineaceae bacterium oral taxon 439 is one of the most abundant taxa on average across the pre-modern populations (Figure 3.5). Its variation is diagnostic of population structure and it has been shown to evolutionarily diverge across geographic and dietary transitions (Eisenhofer et al., 2020; Honap et al., 2023; Ottoni et al., 2021;

Velsko et al., 2024). A high-quality representative MAG was constructed from a Wichita metagenome (WAMP26). Alignment of comparative short-read metagenomes to the MAG reference resulted in 22 draft genomes, including historic (n=7), modern, (n=7), Muwekma Ohlone (n=6), and PM-MRSC (n=3). Although many Wichita metagenomic alignments resulted in at least 30% of the reference covered at five times, none had a heterozygosity ratio < 1. The American strains display geographic clustering within a subtree that is separate from those of historic UK and modern Spain.

The core gene alignment was produced with only high-quality individually assembled MAGs as using both medium- and high-quality MAGs to determine a pangenome resulted in a phylogeny with poor bootstrap support. The only modern published reference used in the phylogeny forms a branch distinct from the MAGs generated in this study. The Muwekma Ohlone strains do not cluster together in this phylogeny but are divided by clusters from other populations. The ancient Muwekma Ohlone strain pre-dating acorn consumption clusters with the pre-agricultural PM-MRSC and Transitional-MRSC. The sole SM-MRSC strain clusters with the other agricultural Wichita strains.

3.5 Discussion

Ancient metagenomes have the potential to unlock previously unrecognized microbial diversity, lost to time due to host lifestyle shifts, epidemiological transitions, and strain evolution. From these data, we can infer ecological dynamics in response to host dietary practice and situate human-associated oral microbiomes within a one health

context. For Ancestors settling the coastal regions of what is now northern California, subsistence practices were composed of hunting and fishing with a heavy reliance on acorns as a dietary staple (Milliken et al., 2007). Up until European colonial occupation, Muwekma Ohlone Ancestors were not traditional agriculturalists but lived in permanent settlements and utilized the available surrounding resources. They promoted the growth of disturbance-following small-seeded plants by strategically burning areas of the landscape and the majority of individuals described in this study had a staple diet of acorns and seeds. We have previously observed that a diet high in starch, such as in this case, is enriched in *Streptococcus* species (Chapter 2) (Fellows Yates (B) et al., 2021). This accompaniment is a product of the co-evolution alongside the increase in copy number of human salivary alpha-amylase genes and the ability of *Streptococcus* species to utilize host amylase for starch consumption (Bolognini et al., 2024; Fellows Yates (B) et al., 2021). The California Ancestors' metagenomes were not dominated by *Streptococcus* species, but were instead enriched in Actinobacteria, such as *Actinomyces dentalis*, *Olsenella* sp. oral taxon 807, and *Pseudopropionibacterium propionicum*, as well as unclassified Firmicutes and *Eubacterium mintutum*. These dominant taxa are shared across comparative populations. *Capnocytophaga*, *Rothia*, and *Neisseria* species are enriched in the modern metagenomes and elevated in the agricultural populations. These taxa are associated with high-fiber diets (Millen et al., 2022) and are a product of the neolithic transition in Europe (Quagliariello et al., 2022). The transition to agriculture was accompanied by a shift in macronutrient proportions and in the types of carbohydrates

consumed (Quagliariello et al., 2022). This likely contributed to microbial composition, particularly with respect to functional capacity.

Though the majority of individuals included in this study were hunters and gatherers, three were recovered from what is now known as the Sanchez Adobe, the site of the former *asistencia*, or sub-mission, called San Pedro y San Pablo which was active in the late 18th century (L. Panich et al., n.d.). The individuals found at this site were thrust into the Spanish mission system and involved in crop production. The difference in dietary habits between this group and the EP and MP had little bearing on taxonomic composition. However, through a principal coordinate analysis of functional composition in the context of other agricultural, hunter-gatherer, and modern data points, we observe a gradient from pre-acorn consumption on the fringes to agricultural practice (Figure 3.7). This pattern could be better understood with greater representation from alternative time points, but it appears that the transition to settlement permanence in the MP, followed by contact with European colonizers, had some effect on microbial functional composition. Further research could investigate whether this observation is due to a dietary influence or is a result of changing population structure, with either settlement permanence leading to greater density or European colonization introducing microbial community exchanges.

Diet affects microbial function with respect to starch and sucrose metabolic functions. Sucrose, an oligosaccharide of starch, is a more modern dietary component (Bai et al., 2017), thus its high abundance in modern calculus is of no surprise. High sucrose consumption prompts the accumulation of fructans and glucans, or polymeric fructose and glucose, due to beta-fructofuranosidase and glycosyltransferase activity

(Buckley et al., 2014; Chakraborty et al., 2022; Millen et al., 2022). The abundance of fructokinase, an enzyme that aids in fructose metabolism, is the most important starch-adjacent enzyme distinguishing populations, and its high abundance corresponds to a high-starch diet (Modern, California, SM-MRSC, and Wichita). Starch synthase, an enzyme that synthesizes and hydrolyzes starch molecules, promotes cell adhesion to tooth surfaces and biofilm formation (Bai et al., 2017) and is abundant in the California and all agricultural populations. It, and other members of glycoside hydrolase families, co-evolved with human dietary shifts towards high starch consumption and the predilection for cooked food (Bai et al., 2017). Starchy foods that are cooked facilitate the amylolytic degradation of glycosidic bonds, allowing for more diverse enzymes facilitating the breakdown of oligosaccharide fragments. These functions emerge from a variety of contributing microorganisms acting in polymicrobial synergy in accordance with observed taxonomic composition, each contributing to the breakdown of starch and its oligosaccharides (Lamont et al., 2018).

The juxtaposition of metagenomic composition and function from dental calculus of California Ancestors against that of other populations throughout history highlights the nuanced dynamics within oral microbial ecologies with respect to human subsistence strategies. The significance of foodways is not only representative of maximizing available resources but is also an indicator of population dynamics and cultural structure. For the Muwekma Ohlone, acorn intensification in the MP coincided with settlement permanence and increases in population sizes (Leventhal et al., 2011) and could have been mediated by inter-group social interactions (Wohlgemuth, 1996). At a microscopic level, acorn

consumption contributed to the abundance of starch and sucrose metabolic functions but did not promote the growth of microbes typically found in agricultural populations. Microbial dynamics are contingent upon population structure and activity, highlighting the connectivity across ecological scales with respect to human activity.

The methods employed above, common in ancient metagenomic studies, seek to characterize community profiles through a reference-based approach. This method is reliant on curated databases largely composed of clinically relevant organisms. There is a hidden layer of population structure at the taxonomic level that is evident in the diversity of species contributing to taxonomic function in this dataset. The stratified functional table outlines species contributing to taxonomic function, but many remain unclassified. Additionally, of the 43 taxa that represent the most abundant species across each population, six lack taxonomic resolution at the genus level. These highlight the microbial diversity hiding in ancient metagenomes. One approach to broaden the repertoire of microbial references is through the creation of metagenomically assembled genomes (MAGs) and standard protocols for establishing species level genome bins (SGBs), or uncultivable operational taxonomic units with an unclassified species designation. Brealey et al. demonstrated that recovery of MAGs from ancient contexts is possible through the *de novo* assembly of non-human oral metagenomes (Brealey et al., 2020). Studies have since used this approach to improve functional annotation of paleofeces metagenomes (Wibowo et al., 2021), investigate microbial response to environmental perturbation and climate change (Perez-Mon et al., 2021; Sipes et al., 2021), elucidate strain separation and fine-tune phylogenetic tree construction (Granehäll et al., 2021; Jackson et al., 2024;

Quagliariello et al., 2022; Velsko et al., 2024), and develop natural products from Neanderthal dental calculus (Klapper et al., 2023). Here, we have reconstructed MAGs from three ancient metagenomic datasets investigating microbial ecological dynamics across the American continent with varying subsistence strategies.

People first populated the Americas tens of thousands of years ago in the last great continental expansion of the human species, carrying their associated microbiomes with them. In the period that followed, the first Americans migrated throughout the continents, adapted to climatic fluctuations, built complex societies, and developed innovative technologies to leverage the resources of their surrounding environments, all while isolated from the rest of the planet. The arrival of Europeans in the 15th century marks the beginning of the Columbian exchange: the irretrievable alteration of production, dissemination, and consumption of food ways (Murphy & Klaus, 2017), the destruction of environmental structures (Garland et al., 2016; Klaus & Alvarez-Calderón, 2017), and devastating demographic consequences for the indigenous peoples (Cook, 1998; Crosby, 1973; Hill, 1998). This contact also brought a bidirectional exchange of pathogens and human-associated microbiota. We have seen previously that microbial strain evolution is not prescriptive. Strains of oral microbes vulnerable to agricultural transitions may persist through time, such as with *Pseudoramibacter alactolyticus* (Chapter 2). Opportunistic pathogens unaffected by subsistence pattern may be completely replaced by non-native strains, as with *Tannerella forsythia* (Chapter 2). Indigenous contact with European populations, the advent of industrialization (Gancz et al., 2023), and the progression towards globalization each offers different selective pressures and an opportunity for the

exchange of microbial strains. What drives these evolutionary patterns among host-associated microbiomes is not well understood, but with this study, we have reconstructed microbial strains from the Americas pre-dating the Columbian exchange to offer a baseline for American microbial diversity. We expanded upon this by investigating evolutionary relationships among an oral microbe frequently discussed in ancient metagenomic analyses: Anaerolineaceae bacterium oral taxon 439 (Honap et al., 2023; Ottoni et al., 2021; Velsko et al., 2024), and some of note that have historically been difficult to reconstruct, *Streptococcus* sp. (Velsko & Warinner, 2025). and *Eubacterium minutum*.

In previous studies, Anaerolineaceae bacterium oral taxon 439 strains have been shown to differentiate metagenomes geographically, but the relatively high rate of heterozygosity in reference-based reconstructions suggests the available reference genome may represent more than one strain or species (Honap et al., 2023; Ottoni et al., 2021; Velsko et al., 2024). Our *de novo* reconstructed core gene alignment does not display the distinctive geographic or temporal structure observed previously. The fact that the published genome forms a lineage distinct from the MAGs generated in this study and the lack of geographic congruence observed within tree topology supports the suggestion that the published isolate represents two closely related strains or species (Velsko et al., 2024). Geographic structuring is observed in the American strains for the MAG reference-based reconstruction. The resolution of the lineage containing those from the UK and Europe is less resolved, but the historic samples lie basal to the modern strains and published isolate. Individual assembly is sufficient for generating high quality MAGs, but

deeper sequencing and more widespread reconstruction of non-American MAGs may help resolve the evolutionary history of this organism.

E. minutum is not well-characterized, but is consistently present at high abundance in ancient oral microbiomes (Figure 3.5) (Kazarina et al., 2021; Putrino et al., 2024; Velsko et al., 2022). Through personal observations, we have found difficulty in resolving tree structure in previous attempts. However, we obtained high-quality MAGs from 33 American samples. The core gene phylogeny has strong bootstrap support and samples sharing temporal period and geographic location generally cluster together. Wichita strains do not follow this pattern and are consistently basal to more branching lineages. The hunter-gatherer populations of the Ancestral Maya and Muwekma Ohlone form a sublineage and are segregated from the Agricultural Maya and Wichita. This suggests a possible dietary influence in the evolution of *E. minutum*. The reference-based phylogeny has two distinct lineages, but bootstrap support weakens after initial branching. Notably, the modern samples sharing a lineage with the historic Radcliffe samples includes those isolates from Europe, China, and the USA. The American lineage does not display the homogeneity of tree placement for California Ancestors observed in the core gene phylogeny and the resolution is poor. However, the two samples that predate intensive acorn consumption cluster together basal to other California samples. Despite the agricultural diet practiced by those in the SP, the strain from this period remains clustered with the other California strains. The dietary signature may be present in this reference-based alignment given the distinct separation between PM- and SM-MRSC samples, though the signal is not as strong. More stringent criteria for inclusion in the phylogeny and

a wider scope for populations included in the alignment may help resolve some of the poor resolution.

The expansion of *Streptococcus* diversity is strongly indicative of agricultural transitions and the modernized diet (Bai et al., 2017; Fellows Yates (B) et al., 2021; Lamont et al., 2018; Millen et al., 2022; Velsko & Warinner, 2025). Its saccharolytic nature thrives in the presence of fermentable carbohydrates and can withstand low pH levels (Abranches et al., 2018; Kianoush et al., 2014). The most abundant *Streptococcus* species in ancient populations is consistently *S. sinensis*, though this discovery has arisen in light of greater diversity in reference databases (Velsko & Warinner, 2025). Modern dental calculus has a greater proportion of *S. sanguinis*, and both the modern and historic populations have a high proportion of *S. oralis* (Supplementary Figure 8Supplementary Figure 8 Proportion of *Streptococcus* species in each population.

). The only *Streptococcus* MAGs generated across these populations were classified as *S. sinensis*, which is reflective of its relative abundance. There were more MAGs generated for a core alignment than were recovered from the reference-based methodology due to high rates of heterozygocity. This indicates the presence of more than one strain in ancient oral metagenomes. It is possible that the assembly protocol employed in this study is incorporating the variant sites distinguishing this hidden diversity and further investigation could elucidate this variability.

In both the SNP and core gene alignments, genomes recovered from the Ancestral Maya and Wichita cluster together with high bootstrap support. The modern published isolates, from USA, UK, China, and Japan, also cluster together. The genomes from Muwekma Ohlone Ancestors are distinct from the other strains, suggesting that agricultural practice may drive the phylogenetic arrangement. These two genomes also show a greater variability between the others in pairwise ANI comparisons (Supplementary Figure 9). (Velsko & Warinner, 2025) describe *S. sinensis* as a VANISH-ing taxon of the normal oral flora. It is Volatile and/or Associated Negatively with Industrialized Societies of

Humans, meaning it is not prevalent in communities with high global market integration, arguably because of toothbrushing. The reconstructed *S. sinensis* genomes are the first from ancient contexts, but further research could elucidate the underlying factor complicating its reconstruction.

In most MAG reference-based phylogenies, there is distinct separation between the eastern and western hemispheres, highlighting the benefit of reconstructing MAGs from ancient American metagenomes. Not all evolutionary trajectories are correlated to dietary practice, and when there does appear to be structuring along these lines, the trend is not shared across all populations. Many MAGs were generated through this assembly protocol, including a large number with no species level designation. *De novo* metagenome assembly and genomic reconstruction could expand our understanding of oral taxonomic diversity in the Americas prior to the Columbian exchange. Further research in this realm could elucidate what drives the persistence or loss of microbial strains, and how host ecological dynamics shape this process.

3.6 Conclusions

In collaboration with and with permission from the Muwekma Ohlone Tribe and Amah Mutsun Tribal Band of Mission San Juan Bautista of the San Francisco Bay area, we present the largest oral metagenomic dataset from the Americas to date. Muwekma Ohlone Ancestors spanning the Early, Middle, and Late periods (~3670 YBP-European contact) were hunters, gatherers, and fishers, and relied on high-starch acorns and seeds as dietary staples. The arrival of Spanish colonizers was accompanied by a disintegration

of Indigenous customs and a transition into forced assimilation and peonage (Arellano et al., 2021). Oral microbiome research is firmly situated within a one health framework, which states that the health of any one of the three systems of life, humans, animals, and the environment, is contingent upon the health of the other two (Huang et al., 2024; Silbergeld, 2019). Oral microbial networks are directly impacted by human diet and lifestyles and reciprocally impact human health, thus offering a window into the connectivity between different ecological scales. Despite the small sample size of the EP and SP era, we see a signature of transition between groups. Functional composition, but not taxonomic composition, of SP oral microbiomes more closely resembles that of other agricultural groups, while the EP samples, pre-dating intensive acorn consumption, reside on the external margin of principal coordinate space. Despite a hunting and gathering lifestyle, MP Muwekma Ohlone had elevated functional potential regarding starch and sucrose metabolism compared to another American hunter-gatherer group (Chapter 2). Particularly, there was evidence for diversity in functions that break down oligosaccharide byproducts of starch metabolism and high amounts of starch synthase, a function of biofilm formation.

From metagenomic datasets across pre-European contact Americas, we recovered metagenomically assembled genomes (MAGs) from individually assembled metagenomes. We obtained genomes with known species designations and some that were unclassified beyond the family level, revealing an under-explored microbial diversity of the Americas. Our pangenome analysis and phylogenetic reconstruction of core genes revealed that *Eubacterium minutum* has a strong dietary signature, with predominately

hunter-gatherer populations forming a lineage separate from agricultural populations. In the first reconstructed *Streptococcus sinensis* genomes from ancient contexts, we reveal that, whether from the source of starch consumption or from some hidden driver of evolutionary divergence, Muwekma Ohlone *S. sinensis* genomes are differentiated from modern and pre-modern agricultural populations. Finally, using a MAG derived from ancient metagenomic assemblies as a reference in draft genome reconstruction and variant calling, we improve bootstrap support for a maximum likelihood tree of Anaerolineaceae bacterium oral taxon 439, a well-studied taxon prevalent in ancient metagenomic datasets.

The factors driving microbial composition and function are multifactorial and vulnerable to host lifestyle. Each microbe has its own evolutionary trajectory and European colonization of the Americas altered not only the infectious pathogenic landscape, but also that of commensal and non-infectious pathogens. Our findings highlight the necessity of expanding microbial references to fully describe ancient oral microbiomes and broaden our understanding of oral microbial diversity.

3.7 Acknowledgements

We would like to thank the Muwekma Ohlone Tribe of the San Francisco Bay Area for their permission and encouragement to pursue knowledge pertaining to the health and biology of their Ancestors. We would also like to thank the Amah Mutsun Tribal Band of Mission San Juan Bautista, the most likely descendent tribe for CA-SMA-71/H (Sanchez Adobe), for their support of this research.

CHAPTER 4 – SHOTGUN METAGENOMES FROM BISON DENTAL CALCULUS CAN OFFER INSIGHTS INTO ECOLOGICAL LANDSCAPES

4.1 Abstract

Non-human mammalian dental calculus is a rich source of ecological data that can inform us regarding the interactions between mammalian host populations and their surrounding environments. Here, we present the first exploration of metagenomic composition of dental calculus from the largest ruminant in the western hemisphere with a history deeply intertwined with human populations: the American Bison (*Bison bison*). Being a keystone species in grassland biodiversity, bison may serve as a strong candidate for reconstruction of past ecological landscapes through assessment of eukaryotic DNA derived from archaeological bison calculus. To determine host, prokaryotic, and other eukaryotic DNA signatures present in dental calculus, we sampled archaeological (n=5, ~12,000 YBP) and modern deceased (n=21) bison for a proof-of-concept investigation. Ancient DNA extracts were partially uracil-DNA-glycosylase treated, and both ancient and modern double-stranded DNA libraries were shotgun sequenced using Illumina technology. Metagenomes were taxonomically analyzed using a customized Kraken2 database. Ancient bison calculus showed poor DNA preservation with negligible recovery of any authentic ancient DNA. Near-complete mitochondrial genomes and a high breadth of coverage of the host reference genome were obtained from most modern bison samples. Modern bison calculus was enriched in oral bacteria such as *Arachnia* and

Actinomyces. Bovine *Babesia* parasites were also detected, but poor breadth of coverage precluded positive identification. We recovered > 95% of the genome of a fungal plant pathogen typically dispersed by bison saliva, *Alternaria alternata*, and > 3% of the genome of a dietary plant species, *Sorghum bicolor*. Overall, this study demonstrated that modern bison calculus is a rich source of endogenous host, microbial, and plant pathogen DNA. However, further innovation is required to recover quality oral metagenomes from archaeological bison calculus.

4.2 Introduction

Calcified plaque biofilms, or dental calculus, can harbor biological material for millennia. Its calcified matrix resists taphonomic disruption and can thus be used to view host-associated oral microbiota through deep time scales (Fellows Yates (B) et al., 2021; Warinner et al., 2014; Weyrich et al., 2017). The calcified matrix may also harbor the host genome (Brealey et al., 2020; Moraitou et al., 2022; Ozga & Ottoni, 2023) and other microremains, such as phytoliths or plant material and starch grains (Akersten et al., 1988; Armitage, 1975; de Oliveira et al., 2021). Numerous studies have made use of the advances in ancient DNA (aDNA) technologies and the accessibility of high-throughput shotgun metagenomic sequencing to analyze ancient human oral microbiomes with respect to dietary subsistence (Chapter 3) (Innocenti et al., 2023; Jacobson et al., 2020; Modi et al., 2023; Weyrich et al., 2017), human history and migration (Fleskes et al., 2024; Honap et al., 2023; Ottoni et al., 2021; Velsko et al., 2024), disease status (Jersie-Christensen et al., 2018), and lifestyle transitions (Chapter 2) (Fellows Yates (B) et al.,

2021; Gancz et al., 2023; Quagliariello et al., 2022; Velsko et al., 2019). Fewer studies have focused on reconstructing non-human oral microbiomes (Brealey et al., 2020, 2021; Fellows Yates (B) et al., 2021; Moraitou et al., 2022; Ottoni et al., 2019; Ozga et al., 2019; Ozga & Ottoni, 2023). This approach, however, is firmly situated within a one health framework, where the health of humans, animals, and the environment individually is reciprocally dependent upon these elements collectively. By characterizing the oral microbiota of our non-human mammalian counterparts, we can look beyond human health and biology and expand the reservoir of oral microbial ecologies across ecological networks. This will further contextualize oral microbiomes within a broader, interconnected one health ecosphere.

Most studies of animal oral microbiomes have used aDNA dental calculus recovered from museum and archaeological specimens due to their ubiquity and contextual detail. The objectives for this type of research have included surveillance of anthropogenic antibiotic use in archaeological bears as a proxy for remote environments (Brealey et al., 2021), to observe the effects of diet and altitude on gorilla oral microbiomes (Moraitou et al., 2022), to compare the genomes of opportunistic oral pathogens from humans and chimpanzees (Ozga et al., 2019), and to observe variation between the oral microbial communities of captive and wild baboons (Ottoni et al., 2019). While the majority of studies investigating non-human oral microbiomes have focused on our closest relatives (i.e. nonhuman primates), only one has reported on the oral microbiome of an herbivorous ungulate (reindeer) (Brealey et al., 2020). Here, we present reconstructed oral

microbiomes from dental calculus of an underexplored, historically significant faunal species that is vulnerable to human activity: the American Bison (*Bison bison)*.

The American bison has a history deeply intertwined with the people on this continent. Descendent from ancestral arctic megafauna that went extinct at the end of the Pleistocene, bison are a keystone species, promoting plant diversity with their feeding, migratory, and wallowing activities (Bergmann et al., 2015; Wendt et al., 2022). It has also been argued that bison are potential dispersers of plant-associated microbial assemblages across the plains (Saleh et al., 2023). Bison are migratory ungulates and early Indigenous American groups across the Plains migrated alongside bison herds through seasonal shifts to find sustenance (Lott, 2002). However, the arrival of Europeans in the 17th century severely disrupted the natural connectivity between Indigenous peoples, their land, and the bison populating the Plains. The westward expansion of settler colonial Europeans and their introduction of corn and wheat crops decimated more than 99% of tall grass prairie, a major food source for bison (Isenberg, 2001; Lott, 2002). This human encroachment into bison ranges fragmented their habitats, irreparably disrupting normal bison migratory patterns, forcing the aggregation of herds. The near extinction of bison in the mid 19th century and subsequent conservation efforts and the confinement of herds never returned bison to the patterns they evolved with (Lott, 2002). Today, the study of bison seasonal migration is limited to ethnographic sources (D. B. Bamforth, 1987; Brink, 2008), ecological models (McMillan et al., 2022; Peck, 2004; Plumb et al., 2009; Polley & Collins, 1984), modern bison migratory and dietary reconstructions (Bergmann et al., 2015; Craine, 2021; Davies et al., 2019; Jorns et al., 2020), and, via mass kill sites, faunal remains

(D. E. Bamforth, 1988; Bement & Carter, 2010; Driver & Maxwell, 2013; Niven & Glenn Hill, 1998; Speth, 2020; Todd et al., 1990; Walde, 2006). Noteworthy is the collection of research regarding the distribution of bison mitochondrial haplogroups following the end of the Pleistocene and introgression with cattle after the population bottleneck of the 19th century (Davies et al., 2019; Douglas et al., 2011; Ovchinnikov & McCann, 2023; Stroupe et al., 2022).

Here, we acquired microbial and eukaryotic genetic material from bison dental calculus to expand the available source of bison biogeography. We collected dental calculus from a small archaeological assemblage to establish a proof-of-concept for ancient metagenome reconstruction. Additionally, in collaboration with a local ranch, we sampled calculus from a collection of modern, deceased bison remains to contextualize the bison oral microbiomes across a temporal continuum. Through shotgun metagenomic sequencing of ancient and modern bison oral metagenomes, we characterize microbial taxonomic composition and investigate shared and differentiated taxa across time and between mammalian species. Furthermore, we quantify endogenous bison DNA accessible via dental calculus. Additionally, using a custom-designed database of prokaryotic, viral, and eukaryotic whole-genomes, as well as all available mitochondria and plastids, we demonstrate the acquisition of ecologically derived organisms associated with this keystone host species. Finally, we identify opportunities for future research endeavors informed by this analysis. This study offers the first investigation of ancient or modern oral microbiomes from bison dental calculus.

4.3 Materials and Methods

4.3.2 Laboratory Procedures

Ancient Cooper Site

Ancient bison teeth, provided by the Oklahoma Archaeological Society (OAS), were excavated from the Cooper site (34HP45) by LBC in 1993-94 (Bement, 1997). This site was an arroyo, or dry, steep-sided gully, with evidence of three kill episodes attributed to Folsom hunters (12,900-12,100 calendar years before present (cal. BP)) (Johnson & Bement, 2009). The site is located in northwest Oklahoma at the intersection of the Great Plains and Osage Plains. It is estimated that each kill was orchestrated at the end of summer or early fall based on seasonality and the presence of only calves, cows, and juvenile mandibles (Bement, 1997). Five disarticulated teeth were sampled for this study.

All sample processing, DNA extraction, and library preparation was performed in the dedicated clean lab at the Laboratories of Molecular Anthropology and Microbiome Research (LMAMR). Sample preparation and DNA extraction was performed in adherence to standard protocol for ancient human calculus (Dabney (A) et al., 2013; Orlando et al., 2021). Tooth surfaces were wiped with 20% bleach solution and total calculus was removed from all teeth. Calculus was subsampled to ~ 10mg and irradiated with a UV crosslinker (VWR). Samples were washed with 0.5M EDTA (pH=8.0) to remove surface contaminants and incubated in a fresh 0.5M EDTA solution with proteinase K (Qiagen) to digest the material. Samples were heated to 37ºC on a Thermomixer F1.5 (Eppendorf) for the first 24 hours to promote enzymatic digestion and were transferred 25ºC for the next

three days. After four total days of digestion, DNA was extracted using a modified silica-based protocol to maximize the retention of short fragments (Ozga et al., 2016) and eluted in EB buffer (Qiagen). An extraction blank was included to monitor contamination.

DNA extracts were partially treated with uracil-DNA-glycosylase to repair most age-related DNA damage but retain cytosine and guanine transitions at terminal ends to quantify misincorporations and authenticate aDNA reads (Briggs et al., 2010; Rohland et al., 2015). Double-stranded shotgun DNA libraries were built using the blunt-end single-tube method (Carøe et al., 2017) with end-repair, followed by Illumina adapter ligation and nick fill-in. A library blank with PCR-grade water was included to monitor for contamination. Quantitative PCR (qPCR) was performed with the pre-indexed libraries to quantify the cycle threshold (Ct) and calculate optimal number of indexing PCR cycles. Libraries were dual-indexed in quadruplet to encourage DNA read complexity and uniquely barcoded using the Kapa HiFi Uracil+ amplification enzyme (Kapa Biosystems). After purifying with Sera-Mag SpeedBeads at a 1:1.8 ratio (Honap et al., 2023; Rohland & Reich, 2012), DNA fragment lengths were quantified with TapeStation (Agiilent), libraries were pooled, and size-selected to 150-500 bp using PippenPrep (Sage Biosystems). The final pool was shotgun sequenced on one paired-end 2 x 150 bp run on the Illumina HiSeq at Admera Health in South Plainfield, NJ.

Modern Clearwater Bison Ranch

Modern bison teeth were collected from multiple bone assemblages across the Clearwater Bison Ranch in Blanchard, Oklahoma. Two to three teeth per eleven individuals were retained due to calculus formation. The cause of death of each individual was not

known, but it was noted that much of the herd suffered from parasitic infection exacerbated by draught. Bison are assumed to have died naturally of either anemia or other illness, predation, or sudden trauma in the chutes. Some skeletons in the assemblage had evidence of inflammatory reaction in the periodontia (Figure 4.1).

Figure 4.1 Photo of modern bison oral pathology.
Image is of maxillary dentition.

Much of the land was intentionally seeded with native plant species, including Little (*Schizachyrium*) and Big (*Andropogon*) Bluestem Grass and Buffalo Grass (*Bouteloua*). The bison also had access to the introduced species, Plains Bluestem grass (*Bothriochloa*), millet (*Sorghum*), and a cattle cube protein supplement made from processed grain by-products.

Sampling and DNA extraction occurred over four batches, each accompanied by an extraction blank to monitor for contamination. Digestion and extraction followed the methodology employed for aDNA recovery described above with some deviations due to the extent of calcification. After the initial wash with 0.5M EDTA (pH=8.0) recommended for ancient dental calculus, digestion began with simultaneous addition of proteinase K (20 mg/mL) and 0.5M EDTA with a 37ºC activation for 24 hours before transferring to a thermomixer at room temperature. After five days of digestion, samples were placed on a 50ºC heat block for 2.5 hours and transferred to room temperature for two more days. Sterile pestles were then used to crush the sample pellets, which were then left for another five days of digestion. After a total of 12 days, DNA was extracted using the modified column-based workflow described in detail in Chapter 1 (Ozga et al., 2016). As modern DNA is not fragmented similar to ancient DNA (Pääbo, 1989), DNA extracts and accompanying extraction blanks were sheared to a desired 300 bp using the Q800R3 Sonicator (QSonica) set to 50% amplification with 30 seconds on/30 seconds off for 14 minutes. DNA was cleaned with a 2X Sera-Mag SpeedBead mixture to remove fragments shorter than 100 bp (Borry et al., 2019; Honap et al., 2023; Rohland & Reich, 2012).

Sheared DNA was prepared into double-stranded libraries for shotgun sequencing. End repair and A-tailing, followed by Illumina adapter ligation, was performed using the KAPA Hyper Prep kit (KAPA Biosciences). Unused indices and adapters were removed with a 1.2X bead clean-up and adapter-ligated reads were quantified with a qualitative PCR (qPCR) as described in Chapter 2. Libraries were dual-indexed and uniquely barcoded using the KAPA High Sensitivity Ready Mix over three replicate reactions to maximize DNA

read complexity and randomly distribute biases derived from PCR reactions. After a 1.5X bead-cleanup, libraries were quantified with the TapeStation (Agilent) and pooled into equimolar amounts. After size selection with PippenPrep (Sage Biosystems), the final pool was shotgun sequenced on one paired-end 2 x 150 bp run on the Illumina HiSeq at Admera Health in South Plainfield, NJ.

4.3.3 Data Analysis

Pre-processing

Data pre-processing is described in detail in Chapter 2. Briefly, paired-end FASTQ files for ancient and modern samples were processed with AdapterRemoval v2.3.3 (Schubert et al., 2016) to filter reads shorter than 30 nucleotides (nt), as well as trim reads of low quality (Phred score below 30) or ambiguous (N) bases. Sequencing adapters were removed and reads overlapping by at least 10 nt were merged. Quality-trimmed reads too long to have overlapping regions were stored in forward and reverse truncated read files. For ancient samples, the merged and truncated merged reads were concatenated and used in subsequent analyses. For modern samples, these and the truncated forward reads were concatenated to retain quality filtered reads that were too long to be merged. PCR duplicates were removed from both ancient and modern samples using fastp v0.23.4 (Chen, 2023). To assess human DNA read content due to sample handling, or spurious alignment, deduplicated reads were aligned to the human genome reference (T2T-CHM13v2) using Bowtie v2.3.5.1 (Langmead & Salzberg, 2012) with default parameters. Reads mapping to the human genome were filtered from the alignment file using SAMtools

v1.9 (Danecek et al., 2021) and unmapped reads were converted to FASTQ files with bam2FastQ from BamUtil v1.0.15 (Jun et al., 2015).

Sequence Filtering and Endogenous Host DNA

The human-filtered FASTQ files were aligned to the bison reference genome assembly (ARS-UCSC_bison1.0) (Oppenheimer et al., 2021) to assess and filter host endogenous DNA using the same workflow described in the previous section. Terminal nucleotide transitions were quantified for both ancient and modern DNA sequences mapping to the bison reference using MapDamage v2.2.1 (Jónsson et al., 2013). Reads not mapping to the bison reference genome were converted to FASTQ and used for subsequent analyses, hereafter referred to as "analysis-ready" or "AR" reads.

Reads not filtered of bison DNA were also mapped to the bison reference mitochondrial genome (NC_012346.1) using the Burrows-Wheeler aligner (BWA) v0.7.17 (Li & Durbin, 2009). Ancient sequences were mapped with the *aln* algorithm optimized for short-read mapping with seeding disabled and an allowance of 10% missing data to account for damage at terminal nucleotides. Modern sequences were mapped with the *mem* algorithm, which accommodates Illumina shotgun reads but more stringently restricts mismapping. Mapped reads were sorted and filtered of reads with poor mapping quality scores ($q < 37$) with SAMtools v1.9 (Danecek et al., 2021) and PCR duplicates were removed with DeDup v0.12.8 (Peltzer et al., 2016). Genome quality, or the percent of the reference covered by DNA sequencing reads and the depths at which genomic regions are covered, was measured using Qualimap v2.2.1 (Okonechnikov et al., 2016). For ancient bison sequences, terminal nucleotide transitions were quantified with MapDamage v2.2.1

(Jónsson et al., 2013) and, once complete, the two terminal nucleotides at the 5` and 3` ends were removed with BamUtil v1.0.15 (Jun et al., 2015). VarScan v2.3.9 (Koboldt et al., 2009) was used to identify nucleotides that vary from the reference mitogenome. Consensus variants were called if there was a minimum read depth of 5 at that position and a minimum of three reads supporting the variant, a minimum base quality of 37, a minimum variant allele frequency of 0.2, a minimum frequency of 0.9 to call a homozygote, a p-value threshold of 1, and variants with >90% support were not ignored.

Taxonomic Profiling

AR reads were taxonomically classified using the exact *k*-mer matching tool, Kraken v2.1.2 (Wood et al., 2019), and a custom database containing all representative genomes of prokaryotes from the Genome Taxonomy Database (GTDB) (Parks et al., 2022), and all representative genomes of fungal, protozoal, and viral species, as well as eukaryotic mitochondrial and plastid sequences from the National Center for Biotechnology Information reference sequence database (NCBI RefSeq) (Brister et al., 2015; O'Leary et al., 2016; Tatusova et al., 2016). The Kraken database was built and run using default parameters (a *k*-mer length of 35 and minimizer length of 31), and reports were re-estimated to the species level with Bracken v2.0 (Lu et al., 2017).

Ancient Authentication and Source Tracking

To assess the source of the microbial community composition and measure extent of any contamination, comparative samples from modern human calculus (Velsko et al., 2019) and plaque (HMP, 2012), prairie soil (Fierer et al., 2013), bovine and Eurasian buffalo rumen and intestine (colon, jejunum, and cecum) (Stewart et al., 2019; Thomas et al.,

2017; Tong et al., 2022), non-human mammalian feces (Milani et al., 2020; Tong et al., 2022), and human skin (Oh et al., 2014) were processed in the same manner described above and used as input for SourceTracker v2.0.1 (Knights et al., 2011). The prairie soil was collected from states across the Great Plains, including Texas, Nebraska, Iowa, Minnesota, North and South Dakota, Oklahoma, and Kansas (Fierer et al., 2013).

To prevent the influence of laboratory, handler, or soil contamination on taxonomic composition, taxa identified in extraction and laboratory blanks and those identified as soil and skin at a greater abundance than comparative oral communities were removed. To further ascertain compositional structure of the metagenomes from this study, a principal coordinate analysis (PCoA) was performed on taxonomic abundance files of ancient and modern bison, the sources used in source tracking, and historic mammalian (Brealey et al., 2020; Ottoni et al., 2019; Ozga & Ottoni, 2023) and modern human (Velsko et al., 2019) calculus. The abundance table was loaded into phyloseq v1.42.0 (McMurdie & Holmes, 2013) in R v4.2.2 and, due to the compositional nature of metagenomic data (Gloor et al., 2017; Quinn et al., 2019), the table was center log-ratio (CLR) transformed with the microbiome package v1.20.0 (Lahti & Shetty, 2012). Euclidean distance was computed, and values were ordinated in a PCoA plot with phyloseq.

To quantify the damage patterns diagnostic of aDNA, ancient bison AR reads were aligned to the most abundant oral microbes in these samples. Reads were aligned using BWA *aln* as described previously for alignment to the genomes of the following organisms: *Actinomyces israelii* DSM 43320, *Actinomyces* sp. oral taxon 897, *Arachnia propionica* (three different assemblies designated *A. propionica*, *A. propionica_B*, and *A. propionica_C*

with GTDB), *Lautropia mirabilis*, *Schaalia odontolytica*, and *Actinobaculum* sp. oral taxon 183 str. F0552. Terminal nucleotide damage was assessed with MapDamage2 (Jónsson et al., 2013).

As described in the following Results section, a large proportion of metagenomic reads were attributed to a eukaryotic source. Specifically, the most abundant eukaryotic taxa identified were *Babesia* species. Additionally, ecologically and economically significant fungal and plant species were identified in relatively high amounts. Thus, to assess the possibility of recovering eukaryotic DNA from bison dental calculus, AR reads from both ancient and CWB bison were aligned to the *Babesia bigemina*, and *B. ovata* whole genomes, *Alternaria alternata*, and *Sorghum bicolor* in the same manner described in the host and human genome alignment.

4.4 Results

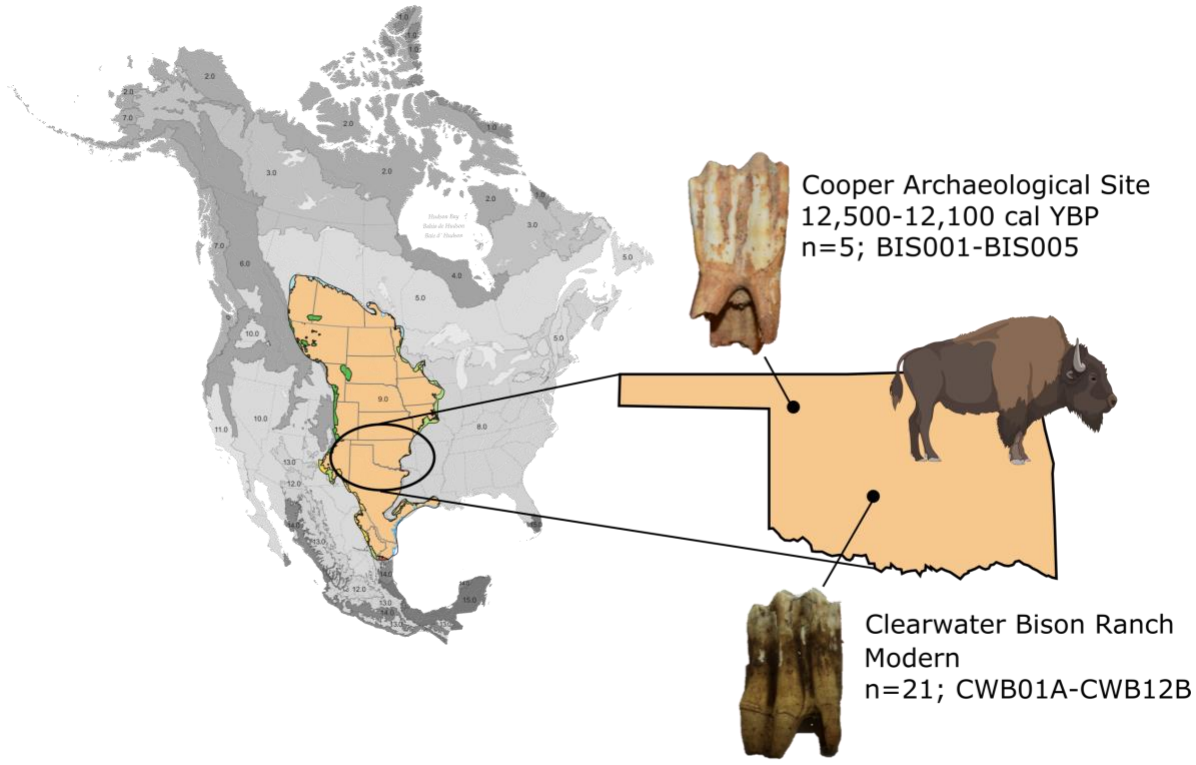


Figure 4.2 Graphical summary.

The section highlighted in the North American map obtained from the Commission for Environmental Cooperation is the extent of the Great Plains region (CEC, 1997). The locations of the archaeological site and modern bison ranch are indicated in the Oklahoma state inset. Bison image retrieved from BioRender (2025) https://app.biorender.com

4.4.1 Shotgun Sequence Processing and Authentication

Raw paired-end shotgun metagenomic data was deposited to the NCBI sequence read archive (SRA) under the BioProject accession PRJNA1258246. The AR reads retained after filtering endogenous bison DNA and possible human DNA contamination ranged from 2,444,000 to 14,849,503 reads for the ancient bison samples and 17,463,553 to 101,061,963 reads for the modern Clearwater bison (CWB) samples. The average read lengths of ancient and modern bison DNA ranged from 60 to 84 nt and 131 to 171 nt,

respectively. Despite decontamination procedures prior to calculus collection and DNA extraction, 47,926 to 888,935 reads mapped to the human reference genome. Detailed results can be found in Supplementary Table 7.

Source tracking was employed to determine the proportion of AR reads attributed to bacterial taxa found in different sources such as human calculus and plaque, cattle and water buffalo stomach (rumen, reticulum, omasum, abomasum) and intestine (cecum, jejunum, colon), mammalian fecal material, prairie soil, and skin. The composition of all ancient and most modern bison samples most closely resembled that of a soil and/or skin microbiome (Figure 4.3). Only two modern bison metagenomes had > 5% reads attributed to taxa found in oral sources. Five modern bison metagenomes had >5% reads attributed to taxa found in ruminant intestinal microbiomes, which might be due to the high genetic identity between bacterial strains found in oral and gut environments which often leads to misclassification. Additionally, large proportions of reads from all bison samples could not be attributed to any known source and were thus classified as “unknown”.

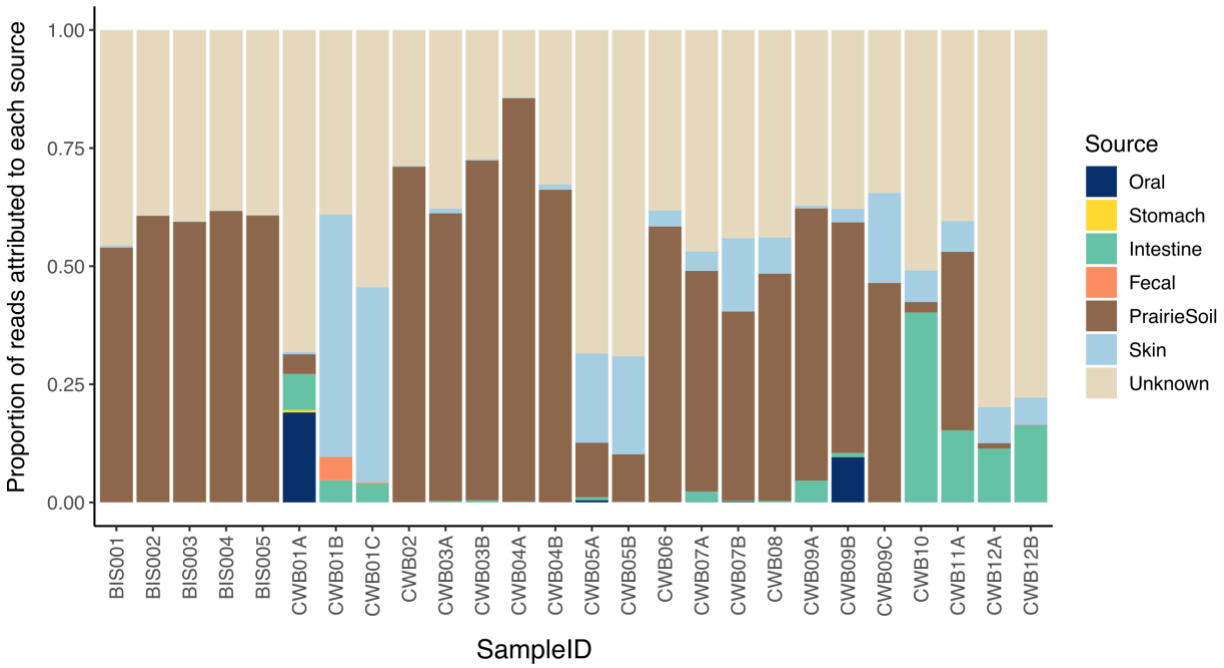


Figure 4.3 Microbial source tracking.

Depiction of proportions of microbial sources composing recovered metagenomes for ancient (BIS) and modern (CWB) bison. Microbial sources indicated by the colors in the legend. Oral sources include modern human calculus and plaque. Taxonomic profiles generated with Kraken2 classification.

Principal coordinate analysis confirmed that both ancient and modern bison metagenomes closely resembled a soil microbiome. In Figure 4.4A, samples from ancient and modern bison cluster in the top left quadrant with the soil samples while the other oral samples cluster centrally. The skin samples are dispersed in the bottom left quadrant. Three samples from a single modern bison individual (CWB01 A, B, and C) cluster with other oral samples. Once soil and skin taxa were removed, modern and ancient bison remain segregated from other populations in the lower left quadrant (Figure 4.4B). Figure 4.4C shows a principal coordinates plot based on only taxa confirmed to originate from an oral source either using the human oral microbiome database (HOMD) (Escapa et al.,

2018) or manually confirming isolate source on GTDB. In this plot, the bison samples cluster together in the top right quadrant with a deer calculus sample.

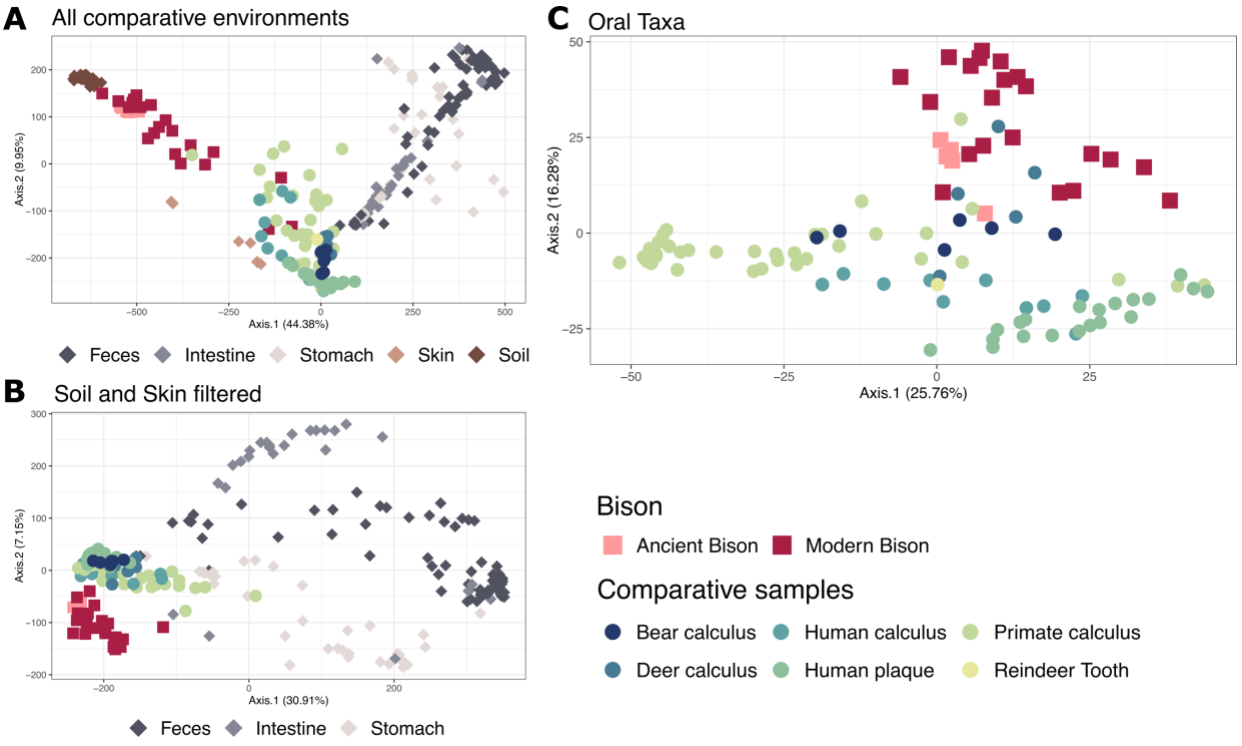


Figure 4.4 PCoA plots of comparative microbiomes.

Relative abundance profiles are filtered of taxa present in blanks and data is CLR-transformed. A) Profile includes soil microbes. Each ancient and modern bison sample is noted. Stomach samples include rumen, reticulum, omasum and abomasum of cattle and water buffalo. Intestine includes cecum, colon, and jejunum. Feces is from many non-human mammalian clades. Skin samples are from a palm swab. B) Plot excludes soil and skin microbes and samples. C) Plot only includes oral taxa and comparative oral samples.

Only one of the ancient bison samples had enough reads aligning to an oral reference genome to quantify aDNA damage (Figure 4.5). BIS001 aligned to five of the most abundant oral taxa with > 1000 reads uniquely mapping to the reference. This sample also had the greatest reads retained after PCR duplicate removal and filtering human and bison DNA reads. Both *Arachnia propionica* strains had the highest number of unique reads mapping to the genome and had elevated frequency of T at the 5` ends, but there is also a

greater frequency of G at those ends. However, the lack of a clear damage signal precludes confident positive identification of the queried oral microbes from ancient bison calculus.

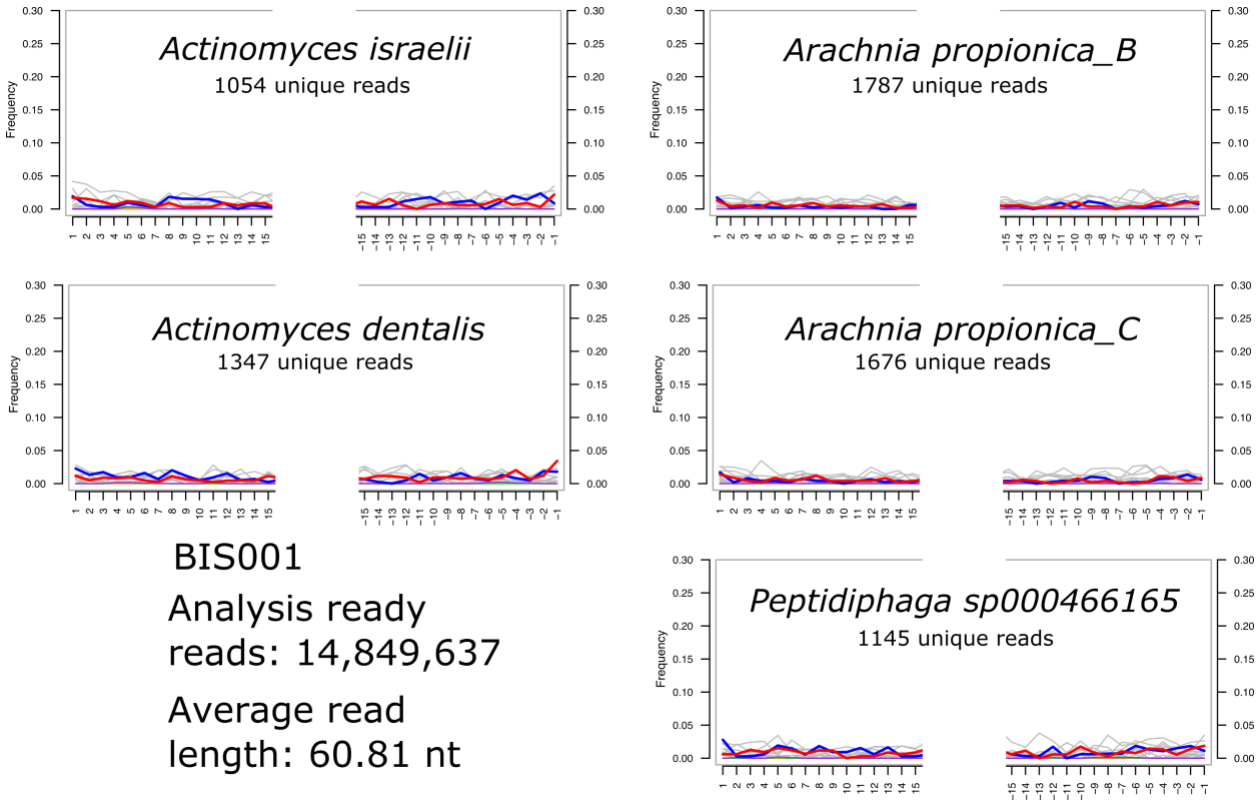


Figure 4.5 aDNA damage plots for BIS001.

MapDamage plots from alignments resulting in >1000 uniquely aligned reads. Red lines indicate thymine (T) and blue lines indicate adenine (A). An upward tick at the left (5`) end indicates a higher rate of T where there is a cytosine (C) in the reference and an upward tick at the right (3`) indicates a higher rate of A where there is a guanine (G) in the reference.

Overall, our analyses suggest relatively low recovery of known oral taxa from the ancient and modern bison calculus samples. The lack of oral taxa and increased presence of soil bacteria in the ancient calculus samples likely indicates poor endogenous DNA preservation. Though the paucity of oral taxa in most of the modern calculus samples may indicate potential sampling issues, there is a possibility that the appropriate oral microbes

to use for ancient DNA authentication are not yet characterized and are thus missing from available databases.

4.4.2 Host Endogenous Content

After removal of contaminant human DNA reads, few reads from the ancient bison samples aligned to the bison reference genome and none aligned to the mitochondrial reference. The modern CWB samples, however, had between 394,583 and 84,473,573 unique reads aligning to the bison reference genome with 1.45% to 95.45% of the reference covered at least one time. In all but four CWB samples, almost 100% of the bison mitochondrial reference was covered at least five times. Overall, our analyses indicate that bison mitochondrial and nuclear DNA can be recovered from dental calculus. With good host genomic coverage, insufficient recovery of oral microbial DNA is not likely due to poor DNA preservation or inefficient extraction methods but is likely a product of including more enamel in the sampling process.

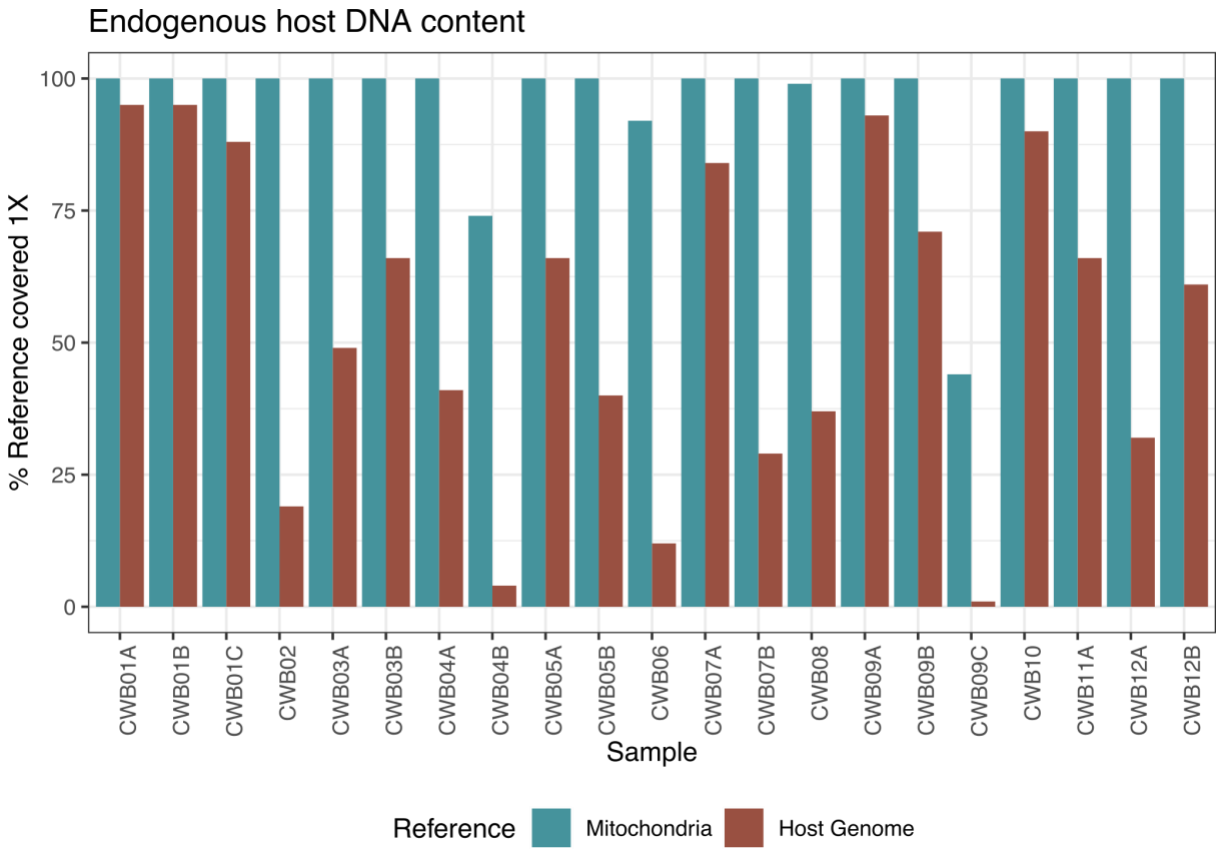


Figure 4.6 Endogenous host DNA.

Proportion of host mitochondrial and nuclear DNA recovered from CWB bison calculus. Too few ancient reads mapped to host genome to be included in this figure.

4.4.3 Taxonomic Composition

Taxonomic profiles provided by Kraken2 included both prokaryotic classifications from GTDB and eukaryotic designations from NCBI from the inclusion of fungal, viral, and protozoan whole genomes, as well as mitochondrial and chloroplast sequences, in the custom Kraken2 database.

Prokaryotic composition

The taxonomic profiles of ancient and modern bison were compared to those of other published human and non-human datasets to observe any patterns across mammalian species. A heatmap depicting the 10 most abundant oral taxa identified in each population (totaling to 43 taxa) is visualized in Figure 4.7. Metagenomes from modern and ancient bison, as well as from bear and some deer but not from humans or nonhuman primates, were enriched in *Arachnia propionica* (sometimes classified as *Pseudopropionibacterium propionicum* (Scholz & Kilian, 2016)). *Actinomyces* species was detected in all calculus but consistently in greater abundance relative to other detected taxa across bison, deer, and bear calculus. *Schaalia odontolytica* (GTDB: *Pauljensenia sp000185285*) and *Actinobaculum* sp. oral taxon 183 str. F0552 (GTDB: *Peptidiphaga sp000466165*) were also detected in ancient and modern bison, and deer, but not in bear oral metagenomes. Though oral microbial recovery was poor, taxa identified in bison after filtering coincide with other oral microbial communities. Particularly, oral microbiota of the only other ruminant calculus sampled, Scandinavian reindeer, resemble those of the bison metagenomes.

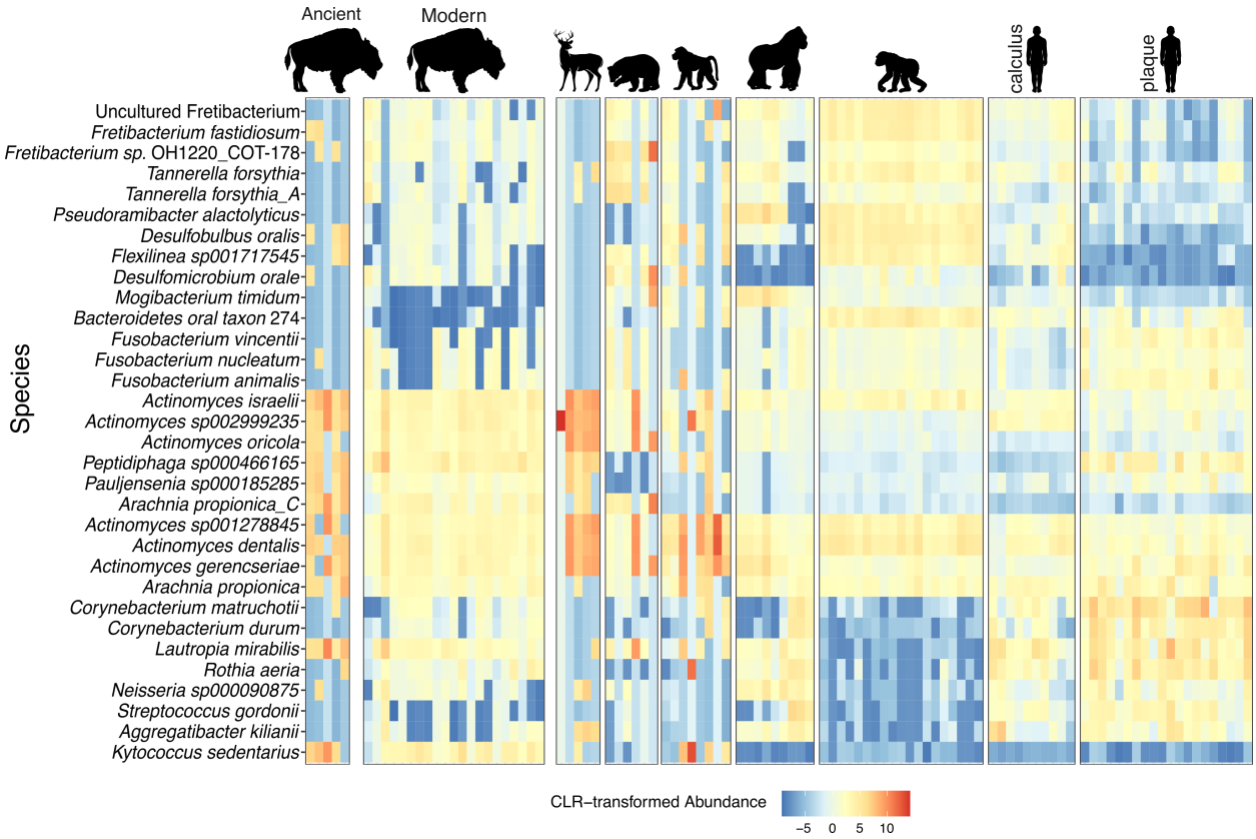


Figure 4.7 Heatmap of comparative calculus populations.

Taxonomic profile generated by kraken2 and filtered to the ten most abundant prokaryotic taxa in each dataset (GTDB nomenclature; total=43 taxa) was CLR-transformed. Samples and taxa were ordered by hierarchical clustering of Euclidean distances using the R base stats v4.2.2 package in R. Comparative populations depicted visually includes reindeer, bears, baboons, gorillas, chimps, and human calculus and plaque.

Eukaryotic composition

Taxonomic profiles of ancient and modern bison, as well as comparative oral and soil samples, were limited to eukaryotic designations to determine the efficacy of acquiring direct detection of dietary or ecological data. The most abundant eukaryotic taxon identified was the Babesiidae family, followed by Ascomycota, though the abundance of the most abundant genus from this phylum, *Alternaria*, is an order of magnitude lower than the *Babesia* genus. We investigated the representation of

Babesiidae across the comparative populations and found an elevated abundance in modern, but not ancient, bison, deer calculus, and bovine and Eurasian buffalo intestinal and rumen samples (Figure 4.8). We isolated the prominent *Babesia* species (*B. bigemina*, *B. bovis*, *B. microti*, *B. ovata*, and *Babesia* sp. Xinjiang) and investigated the relative abundance in each modern bison and soil sample to rule out contamination as the source. The most abundant species across modern samples is *B. ovata*, followed by *B. bigemina* (Figure 4.8B). No *Babesia* species were identified in the prairie soil samples.

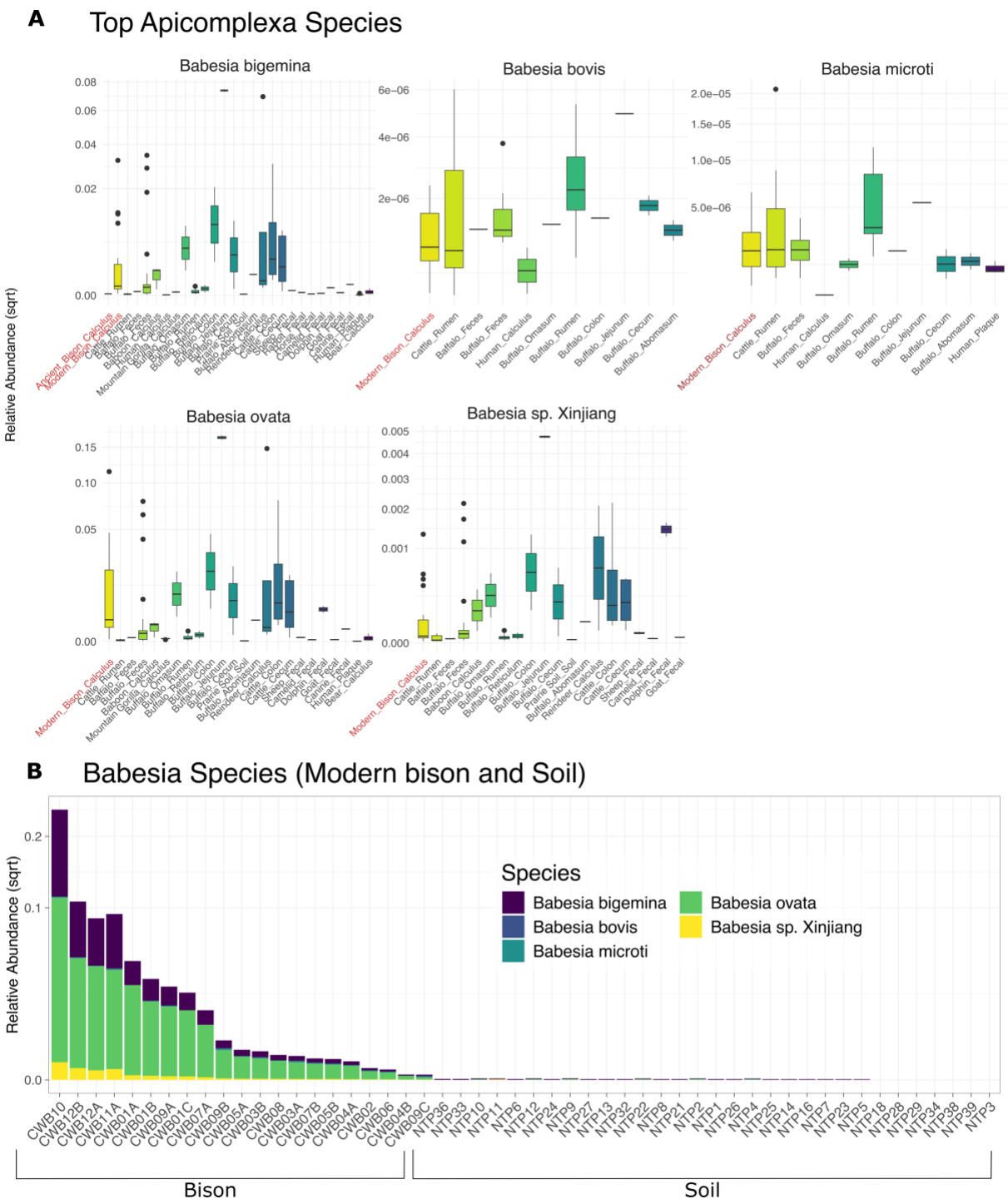


Figure 4.8 Detection of Babesiidae family.

A) Relative abundances of Babesiidae family across comparative populations. Ancient and modern bison are highlighted in red. B) Top five *Babesia* species found in modern CWB samples. Soil samples included to reduce the likelihood that *Babesia* is a soil contaminant.

The second most abundant eukaryotic phylum was Ascomycota. Figure 4.9 outlines the 20 most abundant genera from this phylum in both ancient and modern bison, as well as prairie soil to indicate detection is not due to contamination. We limited our analysis to the 20 most abundant taxa with any reads mapping and found that the most abundant was *Alternaria* species. The abundance of Ascomycota genera in the ancient bison calculus was equivalent or less than that of the soil samples, indicative of a lack of Ascomycota reads in the ancient bison metagenomes.

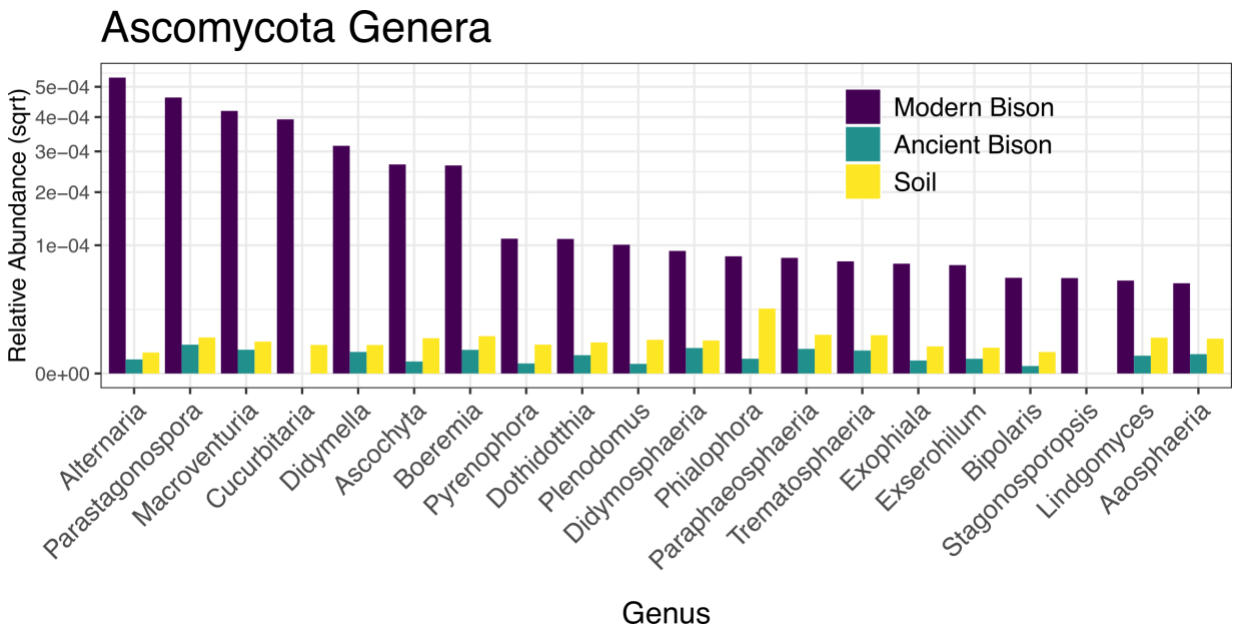


Figure 4.9 Abundant Ascomycota genera.
Bar graph indicating relative abundance in modern, ancient, and soil metagenomes. Y axis is square root-transformed for visual clarity.

Bison are seasonal grazers, subsiding on a variational proportion of grasses, sedges, rushes, legumes, and forbs throughout the year (Bergmann et al., 2015; Craine, 2021). We limited eukaryotic classification to families within these categories and the proportion of reads from samples with detectible amounts are listed in Figure 4.10. The

count of each genus is listed in Table 4.1. CWB01B and CWB05B had the highest count and variety of dietary taxa identified. Only three soil samples had any reads mapping to dietary taxa, and this only included *Eleusine* species. The distribution of grass and legume species are outlined in Figure 4.11.

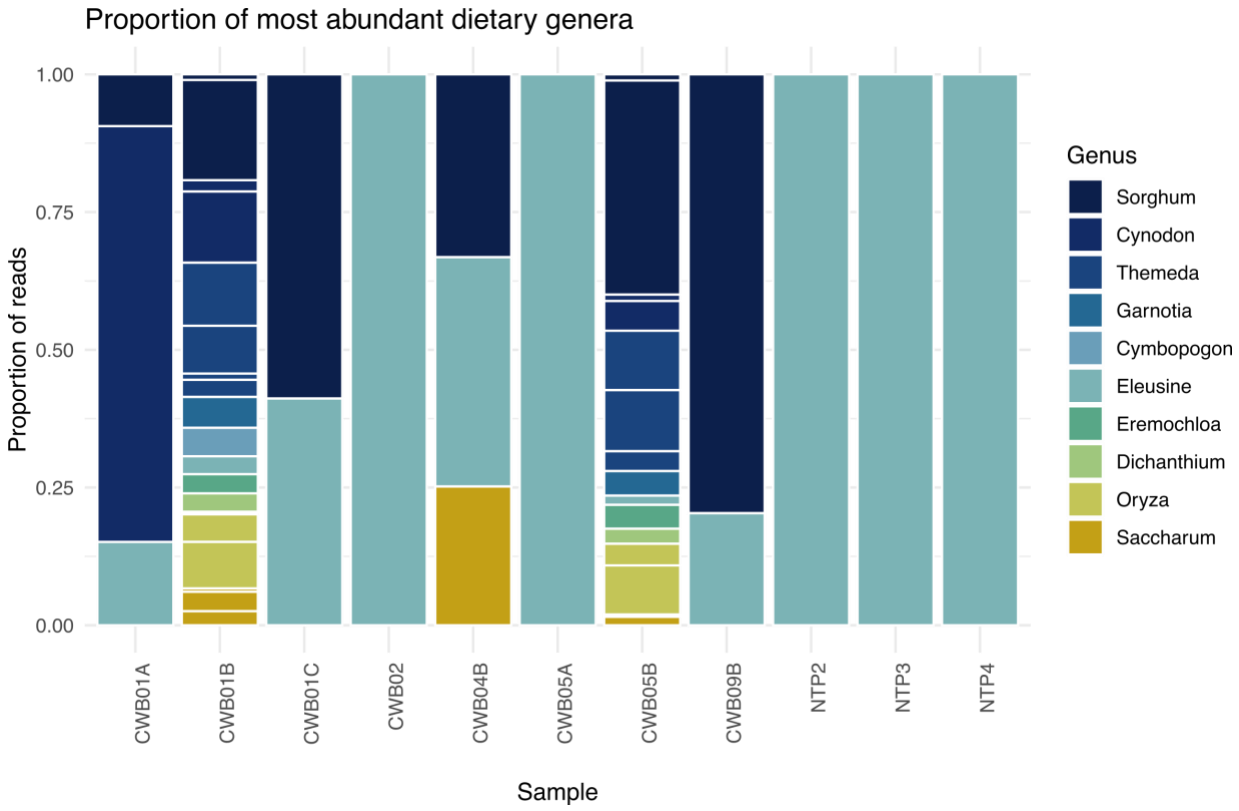


Figure 4.10 Proportion of reads attributed to dietary genera.

Isolated to genera of few forbs, grasses, sedges, rushes, and legumes.

Genus	CWB01A	CWB01B	CWB01C	CWB02	CWB04B	CWB05A	CWB05B	CWB09B	NTP2	NTP3	NTP4
Themeda	0	19446	0	0	0	0	9295	0	0	0	0
Sorghum	105	15345	230	0	137	0	14595	399	0	0	0
Cynodon	843	11969	0	0	0	0	2401	0	0	0	0
Oryza	0	11114	0	0	0	0	4691	0	0	0	0
Saccharum	0	5333	0	0	104	0	710	0	0	0	0
Garnotia	0	4479	0	0	0	0	1630	0	0	0	0

Cymbopogon	0	4138	0	0	0	0	0	0	0	0	0
Eremochloa	0	2778	0	0	0	0	1575	0	0	0	0
Dichanthium	0	2678	0	0	0	0	1003	0	0	0	0
Eleusine	169	2595	161	154	172	227	619	102	107	83	111

Table 4.1 Number of reads classified as dietary genera.

Samples with no reads classified are excluded from the table. No reads from blank samples were attributed to dietary genera.

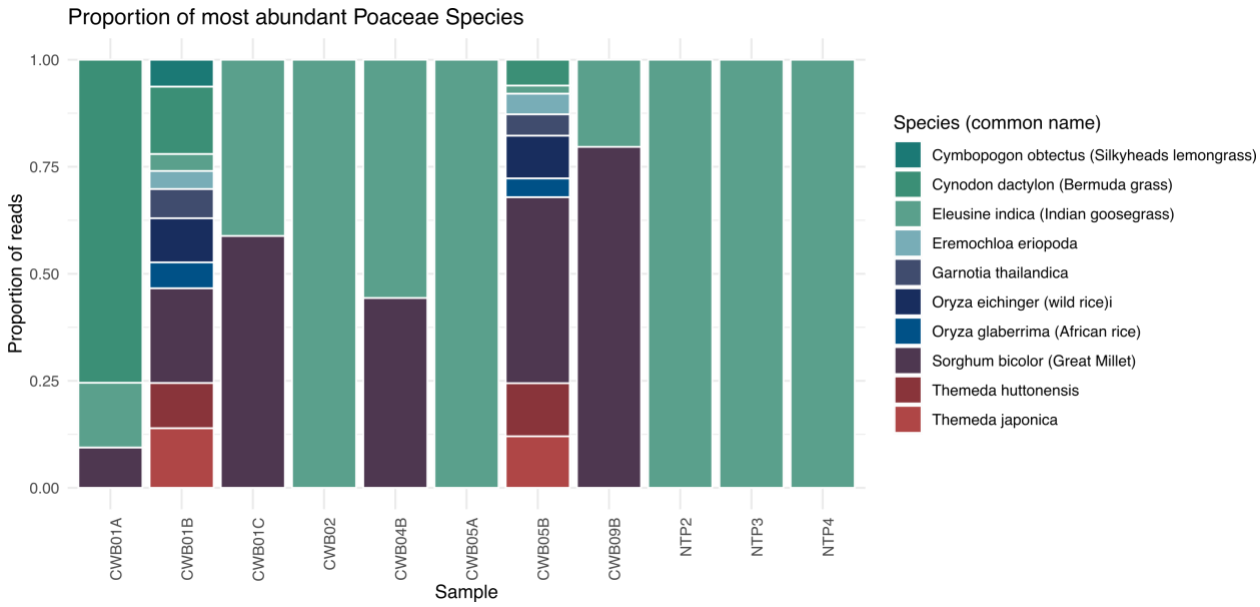


Figure 4.11 Proportion of Poaceae species in CWB and Soil samples.

Samples with no detected amount are excluded. The top ten species are depicted. Common names for species are listed in parenthesis, if applicable.

4.4.4 Eukaryotic DNA recovery

To corroborate the detection of *Babesia* species through taxonomic profiling with Kraken2, we aligned AR reads to the *B. bigemina*, and *B. ovata* whole genomes. These genomes are 13.8 million base pairs (Mbp) and 14.5 Mbp in length, respectively. Each sample had > 2000 unique AR reads mapping to the reference, with an average of 26,784 reads mapping to *B. bigemina* and 32,322 reads mapping to *B. ovata*.

To determine the locations of genomic coverage, deduplicated alignment bam files were assessed with SAMtools *coverage* (Danecek et al., 2021). The output reports the number of AR reads covering each region of the reference. Most alignments to *B. bigemina* were restricted to conserved regions, such as the hypothetical mRNA gene (NW_012237086.1). For other alignments, the distribution of aligned reads is slightly more widespread. CWB11A shows the greatest coverage across both the *B. bigemina* and *B. ovata* references at 0.21% and 0.28%, respectively. For both alignments, the regions with a mean depth of coverage > 1X are conserved, sometimes poorly characterized regions. The region of *B. ovata* with accession number NW_021638399.1 is 8,691 nt long and is 96% covered by CWB11A reads at a depth of 199X. This region is annotated as a putative retrovirus-related Pol poly LINE-1 gene. To visualize the alignment, select deduplicated alignment BAM files from *B. bigemina* were reconfigured into coverage histograms using bedtools (Quinlan & Hall, 2010) and loaded into Circos plots Figure 4.12A (Krzywinski et al., 2009).

		Babesia bigemata				***Babesia ovata***			
Sample	Filtered reads	Unique mapped reads	Mean coverage	% reference covered >= 1X	kraken2 classified reads	Unique mapped reads	Mean coverage	% reference covered >= 1X	kraken2 classified reads
CWB01A	104595906	85957	0.6577X	0.110%	21417	124676	0.9333X	0.250%	300483
CWB01B	97331420	75212	0.5699X	0.110%	17698	109013	0.8065X	0.240%	207997
CWB01C	71053458	52864	0.3929X	0.120%	11974	73653	0.5378X	0.240%	167348
CWB02	90420167	6319	0.0343X	0.170%	853	6555	0.0375X	0.220%	11812
CWB03A	68159308	12430	0.082X	0.190%	2394	16280	0.1112X	0.250%	29615
CWB03B	86117469	19050	0.1288X	0.190%	4258	24475	0.1675X	0.260%	51097
CWB04A	88583181	14837	0.0905X	0.150%	2613	18454	0.1145X	0.220%	33981
CWB04B	72937274	2227	0.0105X	0.130%	199	2156	0.0107X	0.150%	2051
CWB05A	88170742	30775	0.2036X	0.150%	5247	39723	0.2616X	0.250%	66736

CWB05B	90890118	24888	0.1629X	0.180%	3763	27457	0.18X	0.240%	37659
CWB06	74362762	5775	0.0305X	0.170%	598	6350	0.0357X	0.210%	7663
CWB07A	64103648	42880	0.2851X	0.110%	9224	52595	0.3477X	0.220%	121034
CWB07B	65083815	9129	0.0534X	0.160%	1890	11038	0.0673X	0.230%	27109
CWB08	50504050	14441	0.1053X	0.180%	2674	18959	0.138X	0.250%	30210
CWB09A	77558177	45121	0.3496X	0.140%	12568	63717	0.4906X	0.250%	168266
CWB09B	82602568	22859	0.1652X	0.200%	5767	30344	0.2214X	0.270%	76930
CWB09C	101460411	3032	0.0126X	0.200%	315	2456	0.0098X	0.160%	831
CWB10	85237220	37943	0.2402X	0.150%	893772	51659	0.3248X	0.240%	3200170
CWB11A	96897680	27832	0.1902X	0.210%	417088	37570	0.2583X	0.280%	1395435
CWB12A	32821462	9701	0.0558X	0.120%	151779	11869	0.0688X	0.200%	728183
CWB12B	54898298	19199	0.1434X	0.110%	345036	26443	0.1976X	0.220%	1349902

Table 4.2 Mapping statistics to *Babesia* sp.

Table includes the unique number of reads mapping to each genome, as well as the mean coverage and percent of the reference covered at least one time, determined by QualiMap (Okonechnikov et al., 2016). Also indicated is the number of reads assigned to either species by Kraken2.

Another organism identified by Kraken2 to be present in high abundance is *Alternaria alternata*, a saprophytic environmental fungus, the reference genome of which is 33 Mbp long. Three genomes were recovered from CWB calculus at > 95% of the reference covered at least one time with a mean depth of coverage ranging from 4 to 10X. For the high-coverage genomes, the regions with coverage < 50% are between 1,000 and 12,000 nt in length and annotated as hypothetical proteins. The chromosomal scaffolds > 100K nt in length are covered at > 85%. The three alignments with the greatest coverage and the one with the lowest mean depth of coverage are depicted in Figure 4.12B.

	Alternaria alternata		
Sample	Unique mapped reads	Mean coverage	% reference covered >= 1X
CWB01A	36407	0.188X	16.42%
CWB01B	19398	0.0745X	6.90%
CWB01C	7204	0.0329X	2.98%
CWB02	11457	0.0382X	3.14%
CWB03A	24680	0.0758X	4.18%

CWB03B	56203	0.173X	4.45%
CWB04A	3009	0.009X	0.69%
CWB04B	4123	0.0124X	0.84%
CWB05A	34768	0.1448X	11.67%
CWB05B	148172	0.6839X	44.69%
CWB06	56917	0.2166X	12.98%
CWB07A	1791	0.0048X	0.35%
CWB07B	72988	0.3309X	24.77%
CWB08	998423	4.8741X	95.20%
CWB09A	2571	0.0065X	0.42%
CWB09B	19818	0.0761X	5.93%
CWB09C	38193	0.1433X	10.46%
CWB10	1770864	7.5663X	97.37%
CWB11A	2200054	10.1861X	96.86%
CWB12A	18308	0.0701X	5.95%
CWB12B	3235	0.0114X	0.87%

Table 4.3 Mapping statistics to *A. alternata*.

Table includes the unique number of reads mapping to each genome, as well as the mean coverage and percent of the reference covered at least one time, determined by QualiMap (Okonechnikov et al., 2016). Also indicated is the number of reads assigned to either species by Kraken2.

One relevant dietary constituent identified at high abundance in six samples was *Sorghum bicolor*. Alignment to this 708 Mbp genome (cultivar BTx623) resulted in two samples covering > 3% of the genome at least one time with a mean depth of 0.035X and 0.0439X. 225,153 unique reads from CWB01B and 315,493 from CWB05B mapped to this genome, and coverage analysis indicated relatively sporadic areas of alignment Figure 4.12C. The regions with the greatest breadth and depth of coverage were the mitochondria and chloroplast genomes, and nuclear genomes have the greatest coverage at chromosomal ends.

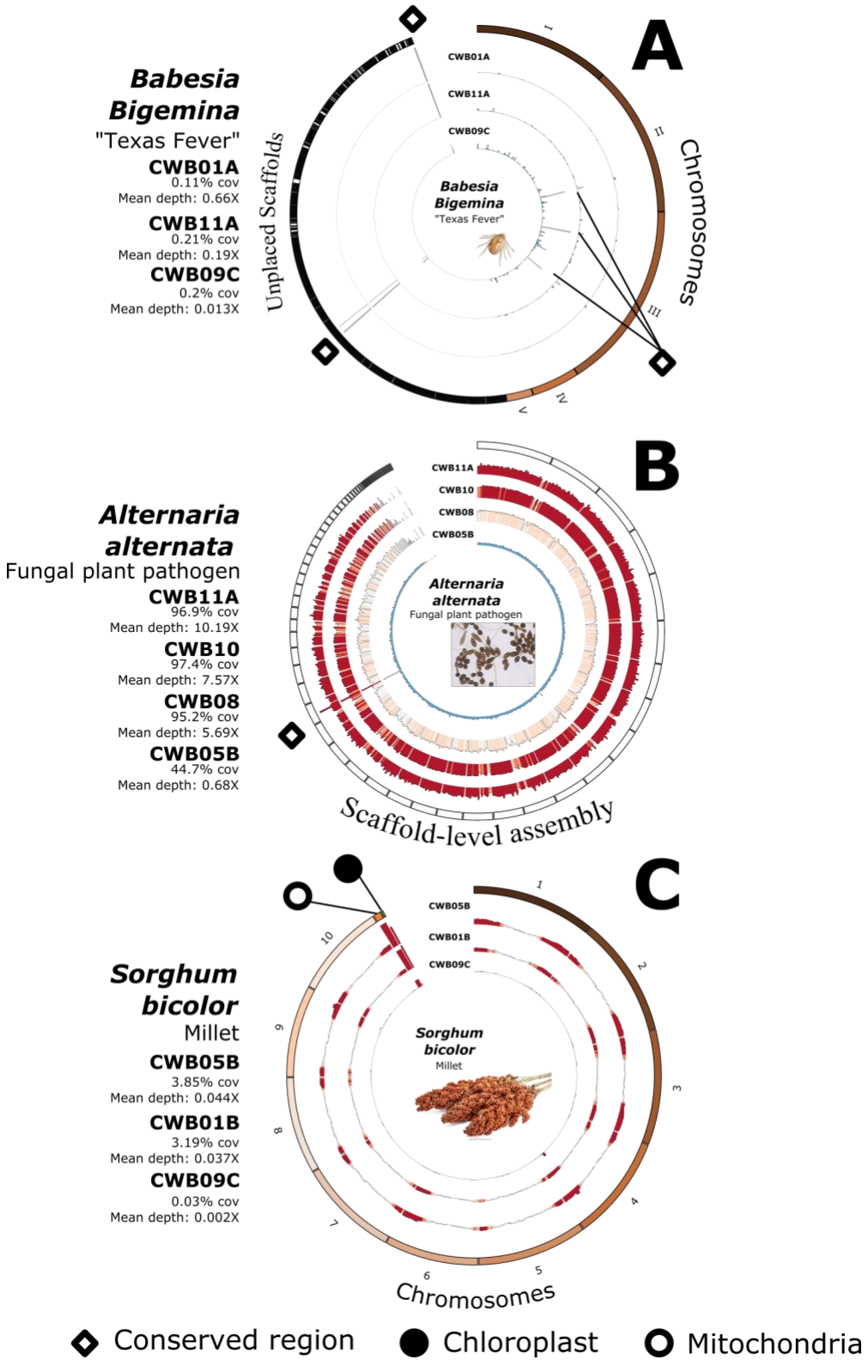


Figure 4.12 Mapping coverage distribution to notable eukaryotic species.

Plots generated using Circos. The outside circle indicates the reference genome at scaffold or chromosome level. The darker the red color on the concentric colors, the greater the depth of coverage. Included genomes are listed to the left of each plot, moving down from most external genome to most internal. A) Coverage distribution for *B. bigemina*. Conserved regions are indicated with the diamond shape. The reference genome is assembled at the chromosome level with hundreds of unplaced contigs. B) Coverage distribution across the *A. alternata* reference. The region with the deepest coverage across all four alignments is indicated as a conserved sequence with the diamond. This reference assembly is at the scaffold level. C) Coverage distribution across *S. bicolor*. Only chromosomal, mitochondrial, and chloroplast regions of the reference are included. The chloroplast is depicted with the closed circle and mitochondria with the open circle.

Despite detection of *B. bigemina* through Kraken2, whole genome alignment revealed a preference for conserved regions, indicating a false identification. However, a fungal plant pathogen was recovered from dental calculus with near complete coverage and high depth in three samples, and eukaryotic DNA from *S. bicolor* was detected with relatively high abundance and even coverage of the reference in two samples, indicating the possibility of recovery of ecologically significant eukaryotes from bison dental calculus.

4.5 Discussion

The North American bison is deeply tied to this continent's ecological identity. The history of this mammal is intertwined with that of human populations of this land, from the first hunters tens of thousands of years ago, through to the near extinction, either directly or indirectly, at the hands of European settlers in the 19th century (Isenberg, 2001), to efforts of revitalization and conservation today. Bison which once roamed from south-central Canada down into central Texas and from the Rocky Mountains to the Mississippi

River (Lott, 2002; Wendt et al., 2022) are now largely limited to private herds relegated to confined acreage (Simpson & Hilken, 2024) due to settler colonial encroachment into North American ecospheres. This reduction of geographic range and introgression with domestic cattle has changed the landscape of bison populations today (Douglas et al., 2011; Stroupe et al., 2022). Furthermore, this keystone species is vital for grassland prairie biodiversity and the impact of their absence has not yet been fully realized, though reintroduction of large herds has had a profound effect on native plant species richness (Ratajczak et al., 2022). Bison reside at the intersection between human populations and North American ecologies, providing a window through which to view one health dynamics of the American Plains. Additionally, bison sit at the forefront of settler colonial activity with the arrival of Europeans. Not only were their rangelands diminished by settlements, but their near extinction was encouraged to intentionally disrupt the connection Indigenous people held with their land and the inhabiting bison populations with the intent to force these displaced peoples onto reservations (Isenberg, 2001; Taschereau Mamers, 2019). The colonial mentality has largely persisted into modernity through the appropriation of bison as a symbol for the American frontier and the commercialization of conservation efforts (Isenberg, 2001). With bison at the intersection between human populations and North American ecologies, an understanding of their oral microbial communities can offer a window through which to view these past biological networks.

Modern bison calculus samples are an excellent source for host endogenous DNA. We recovered nearly complete bison mitochondrial genomes from almost all specimens and more than 50% of the whole bison reference genome from half of the specimens. The

reference genome used in the alignment was an assembly generated form a male F1 hybrid of a male bison – female Simmental cattle cross (ARS-UCSC_bison1 (Oppenheimer et al., 2021). This assembly is represented as more complete than the current available reference genome as it harnesses the variation between the maternal- and paternal-derived haplotype genomes to segregate the haplotypes with more clarity rather than collapse the heterozygosity (Oppenheimer et al., 2021). Unfortunately, for the ancient bison samples, not enough reads mapped to either of the complete bison reference genomes or to the bison mitochondrial genome. Hence, no authentic ancient bison DNA could be detected in the ancient dental calculus samples included in this study. The extensive host endogenous recovery from calculus could indicate that bison enamel or dentin was included during retrieval of the calculus. The poor recovery of known oral microbes described in the next section reiterates this possibility. Further modification of sampling techniques from herbivores could improve the recovery of calculus rather than host bone samples. However, this endeavor showcases the utility of the external portions of bison teeth as an adequate source for bison genomic DNA, at least in modern contexts.

We detected oral microbes from both modern and ancient bison dental calculus, but the former was reduced in recoverable taxa and the latter could not be authenticated as ancient. For both temporal groups, soil contamination appeared to be extensive. However, the inclusion of comparative prairie soil metagenomes in our analysis allowed for the removal of potential soil contaminants. Ungulate dental calculus formation differs drastically from that of humans and non-human primates. Its formation occurs along the surface and within ridges of the teeth (Armitage, 1975). Extensive calcification renders its

removal difficult, and samples require considerable time for chemical digestion. Outside of a resemblance to soil and skin metagenomes, a large proportion of both ancient and modern metagenomes could not be attributed to a source given our available reference databases (Figure 4.3). This is despite the inclusion of bovine stomach, intestine, and fecal metagenomes as sources for the SourceTracker analysis. When the microbial profiles were limited to oral taxa, ancient and modern bison overlapped with the other ruminant (reindeer) dental calculus. Historic reindeer and all bison also show a similar profile regarding abundant oral microbes, with enrichment in *Arachnia* and *Actinomyces* bacteria (Figure 4.7). With the exception of one sample (BIS001), ancient bison metagenomes were unable to be authenticated as ancient due to too few reads mapping to oral reference genomes. The damage profile for BIS001 was not sufficient for ancient DNA authentication. However, this may be due to the lack of adequate bison oral microbial representation in known databases. More extensive surveying of mammalian oral microbial ecologies could elucidate hidden species diversity among bison.

Modern bison calculus is a rich source of microbial and eukaryotic DNA. Some of the jaw bones at the bone piles had evidence of periodontitis (Grzeczka et al., 2023; Jansen et al., 2022) and bone resorption (Figure 4.1). In ruminants, this can be associated *Fusobacterium*, *Synergistetes* (GTDB: CAJPSE01 and *Fretibacterium* in Figure 4.7), and *Elusimicrobia* (GTDB: *Elusimicrobiales*) (Grzeczka et al., 2023), all of which were identified in the modern specimens. *Actinomyces* bacteria, found in both ancient and modern bison, are associated with cara inchada (“swollen face”), which is marked by a rapid progression of periodontal disease (Döbereiner et al., 2000). However, these bison did not have the

periodontal lesions that accompany this disease. Notably, compared to other non-human oral microbiomes, bison have an elevated abundance of *Rothia aeria* and *Neisseria elongata*, nitrate-reducing bacteria associated periodontal health in human populations (Mazurel et al., 2023). This could be correlated to the high nitrogen content in bison saliva (Simpson & Hilken, 2024). It is important to note that the presence of taxa associated with disease is not indicative of illness from that infection.

The most abundant eukaryotes detected in bison calculus includes bovine parasites of the Apicomplexa phylum. *Babesia bigemina* is a tick-borne parasite that can cause babesiosis, or "Texas Fever", in large domestic ruminants and bison (Bock et al., 2004). This parasitic disease was eradicated from domestic cattle in the mid 20th century in the USA by eliminating the tick vector, *Boophilus annulatus* (Zaugg & Kuttler, 1987). However, distribution of this and other vectors is still widespread in Central and South America. *B. divergens* and an accompanying parasite, *Theileria* sp. was detected in American bison in Romania within the past five years (Corduneanu et al., 2023). *B. bigemina* and *B. ovata* were both detected by Kraken2 in the modern CWB bison at relatively high rates. However, direct alignment to the reference genomes resulted in poor distribution of coverage across the genome. A greater proportion of reads were aligned to more genomic regions of *B. ovata*, a parasite of the same genus that can cause inflammatory disease but is not associated with severe symptoms (Yamagishi et al., 2017). This parasite is more widespread in eastern Asia and is transmitted via the *Haemaphysalis longicornis* tick. The only regions of *B. bigemina* with read alignment are highly conserved hypothetical mRNA genes that, when compared to the non-redundant nucleotide

database using BLASTn (Altschul et al., 1990), maps to *B. ovata* and many species of *Bos*. Query sequences mapped more evenly across the *B. ovata* genome, but were primarily localized to the conserved regions, putative retrovirus-related Pol poly LINE-1 and endonuclease-reverse transcriptase. These are transposable elements ubiquitous in parasitic genomes that aid in their pathogenicity (Rodriguez & Makalowski, 2022). Their presence in *Babesia* species renders annotation difficult, resulting in many "hypothetical proteins" and may have contributed to poor DNA recovery through traditional read-mapping performed here. Thus, we cannot conclusively authenticate the detection of *B. bigemina* or *B. ovata* in the bison calculus. Mann et al. provide guidelines for identifying dietary constituents in ancient dental calculus that could be modified here for eukaryotic identification generally speaking (Mann et al., 2023). A *Babesia* signature was identified in all samples, but not any of the blanks or comparative prairie soil samples, so its introduction is not from laboratory contamination and, though there was no soil collection at the ranch, this taxon is not readily found in prairie soil. When comparing relative abundance of species form this taxon across comparative datasets, its prevalence and abundance remains elevated in bovine feces, stomach, and intestine and reindeer calculus, but not other non-ruminant oral environments (Figure 4.8A). The database used in initial taxonomic screening with Kraken2 was complex and included representative genomes of species rather than the entire non-redundant RefSeq database. Although this complexity is designed to promote competitive mapping to deter spurious alignments to highly represented genomes, there is still a possibility of mismapping with a k-mer based approach if taxa are lacking representation (Sczyrba et al., 2017). The list of suggestions

provided in (Mann et al., 2023) includes authentication of aDNA, which is not applicable for modern bison calculus. Another suggestion is to use intuition when interpreting results. For instance, a parasitic signature would "make sense" for this bison herd as individuals have been suffering from anemic disease exacerbated by draught, but the eradication of *B. bigemina* indicates an alternative etiology for such a clinical manifestation. Given the comparative results outlined in Figure 4.8 and the detection of endonucleases, it is possible that another parasitic organism that plagues ruminant species is present, but the species designation cannot be determined with present techniques. Many of the calculus samples had a high host endogenous content, suggesting a possible immune response to a pathogen (Austin et al., 2022; Warinner et al., 2014) and increasing the likelihood of detecting a blood-based parasitic organism in calculus. Of the 65 species listed as *Babesia* in NCBI, only 10 have published reference genomes. Additionally, the genomes used for reference-based mapping were recovered from cattle blood. The detection of conserved regions also identified as cattle using BLASTn draws into question the purity of the reference genome. This region shared between a bovine parasite and bovine host species could indicate an error in the reference genome (Mann et al., 2023; Merchant et al., 2014; Orosz, 2015; Steinegger & Salzberg, 2020). This latter case was echoed by an epidemiologist for the Cattle Fever Tick Eradication Program of the USDA through personal correspondence as the most likely scenario. Though "Texas fever" from *Babesia* species was eradicated from this region, the United States imports many cattle from Mexico, where the disease is still endemic (Santamaria et al., 2024). There are rising concerns for increased incidence of *Babesia* infection in U.S. cattle, but the isolation of the bison herd

at the Clearwater ranch renders their acquisition of this parasite unlikely. Effective confirmation of *Babesia* in calculus requires a more vigorous approach, such as DNA capture for unique, identifying genomic regions (Maixner et al., 2018; Sawafuji et al., 2020), more widespread genomic sequencing of parasitic species to expand the available reference database, or an investigation into possible errors in the reference genome (Orosz, 2015; Steinegger & Salzberg, 2020).

Bison are known to harbor fungal plant pathogens in their saliva and have the potential to act as dispersers for these microbes (Saleh et al., 2023). One such pathogen was identified in high abundance in most specimens: *Alternaria alternata*. DeMers describes this species as “cosmopolitan”, having a history wrought with inconsistent nomenclature and taxonomy (DeMers, 2022). It has a wide host range and can either be a symbiont or pathogen to its plant host. We were able to recover > 95% of the 33 Mb reference genome with relatively deep coverage from three of the modern bison teeth. Unfortunately, it is not possible to elucidate the origin of this recovered genome. All the bison teeth were from completely desiccated remains, estimated to have died within the past five years. *A. alternata* could have been introduced post-mortem, but not in an indiscriminate fashion. With its detection at such a high rate in only three of the individuals irrespective of sequencing depth and its absence in the blanks, it could have been incorporated into the calculus of the individuals.

We were able to detect eukaryotic DNA associated with diet. The lack of whole plant genomes in the reference Kraken2 database reduced the sensitivity of detection, precluding complete dietary reconstruction of bison from calculus (Moraitou et al., 2022;

Ottoni et al., 2019), but excluding these genomes rendered the analysis less computationally demanding and screening provided a basis upon which to recover dietary reads through reference-based alignment. Through Kraken2 screening, we detected signatures of bluestem grasses (*Schizachyrium*) and buffalo grass (*Bouteloua*) in low amounts in two samples, but the most abundant dietary constituent identified was *Sorghum bicolor*, a common component of animal feed. CWB01B and CWB05B had > 200,000 unique reads mapping to the *S. bicolor* reference. This mapping was evenly distributed at the chromosomal ends across the genome, covering > 3% of the 400 Mbp genome. Dietary constituents are rarely found in ancient dental calculus (Brealey et al., 2020; Ottoni et al., 2019; Ozga & Ottoni, 2023). Eukaryotic genomes are much larger and littered with repetitive sequences and representative sequences for dietary sources are lacking in available databases (Mann et al., 2023). Given the difficulty in validating alignments where very few reads align, it is sometimes impossible to verify that reads classified as eukaryotes are true classifications. In historic contexts, typical methods of validation using aDNA damage patterns is difficult when so few reads align to the reference (Brealey et al., 2020). However, Brealey et al. had more success recovering dietary eukaryotic DNA from historic reindeer calculus than either gorilla or bear. This could be due to the exclusive plant-based diet, the material of which is more readily detected in dental calculus (Moraitou et al., 2022). Also, it could be the sheer quantity of food consumed by grazing animals compared to the foraging practice of bears and primates. The *S. bicolor* regions with the greatest depth of coverage from the alignment were the mitochondria and chloroplast genomes. This highlights the success observed from the

custom Kraken2 database used in this study. Although there is a loss of sensitivity in the detection of dietary constituents when whole genomes are not included in the reference database, the use of mitochondrial and chloroplast genomes is an effective alternative for preliminary screening. Ultimately, bison calculus offers an effective medium for detecting ecologically derived genetic material.

4.6 Conclusions

This study marks the first exploration into the recovery of metagenomic information from bison dental calculus. We found that it is possible to retrieve oral-associated microorganisms from this ruminant species, though strict decontamination protocols must be adhered to. Bison oral microbiomes are enriched in *Actinomyces* and *Arachnia* species and have a high abundance of the nitrate-reducing *Rothia aeria* and *Neisseria elongata* associated with periodontal health. Bison calculus also harbors non-prokaryotic DNA. In addition to the recovery of near complete host mitochondrial and genomic DNA, we detected markers for bovine *Babesia* parasites. However, further innovation is required to validate this finding. We were more successful in recovering an even distribution of DNA for genomes associated with the interface between bison and their ecological environment. Bison are fungal dispersers and a keystone species with respect to grassland ecology, and we were able to identify a putative fungal plant pathogen, *Alternaria alternata*, and a commercially significant grass species, *Sorghum bicolor*. Further research could improve validation of these discoveries, but this study demonstrates that bison are good candidates to further innovate methods of dietary reconstruction from dental calculus

metagenomes. Recovery of oral microbial ecologies could not be authenticated from 12,000-year-old bison calculus. However, the inclusion of a modern counterpart indicates a proof-of-concept for the possible successful reconstruction in the future. Using ancient DNA, we could look deep in time at the ecological context of past bison herds prior to European expansion and investigate oral microbial ecologies of the largest ruminant in the Americas with a history deeply intertwined with human populations.

Acknowledgements

We would like to thank the Oklahoma Archaeological Survey (OAS) for their contribution of ancient bison teeth and their guidance through methods development. We would also like to thank Gary Bell and the Clearwater Bison Ranch for their openness for scientific advancement and expertise in bison diet, habits, and health patterns.

CHAPTER 5 – CONCLUSIONS

This dissertation highlights the complexity at the interface between diet, microbial ecology, and health. With each chapter, we interpret microbial compositional patterns through metabolic dynamics within a one health framework. In looking at the organizational system of microbiomes within dietary structures, we demonstrate how molecular anthropological techniques can facilitate our understanding of microbial metabolic mechanisms and elucidate the impact of human activity on microbial evolutionary trajectories.

In chapter 2, "Shifting oral microbial ecologies from metagenomes accompany dietary transitions among ancestral Maya in Belize", we demonstrate that the transition to maize agriculture was accompanied by a decrease in abundance of Chlostridiales bacteria, such as *Pseudoramibacter alactolyticus* and *Eubacterium minutum*, and an increase in saccharolytic *Streptococcus* and *Actinomyces* species. This was likely due to the availability of fermentable carbohydrates and the low capacity for acid tolerance in *P. alactolyticus* and *E. minutum* compared to saccharolytic organisms. Additionally, we found that, though the dietary shift influenced the differential abundance of some taxa, evolutionary mechanisms of bacterial strain diversity are more complex. Though reduced after the adoption of maize, American *P. alactolyticus* has persisted through colonial expansion and modernization, but the opportunistic pathogen, *Tannerella forsythia*, has been completely replaced since the arrival of Europeans. This chapter marks the first

investigation into the impact of an agricultural transition on the oral microbiome in the Americas.

Chapter 3, titled “Oral metagenomes of Muwekma Ohlone Ancestors in California offer insights into pre-Columbian microbial strain diversity and ecological dynamics”, presents the largest oral metagenomic dataset in the Americas to date. Despite a hunting and gathering lifestyle, intensive acorn consumption among the Muwekma Ohlone Ancestors of the Middle and Late Periods (2,450 YBP-European contact) was marked by an increase in functions related to starch and sucrose metabolism compared to the pre-maize Ancestral Maya but were not enriched in *Streptococcus* species like other agricultural groups. Additionally, with respect to microbial functional composition, there is a signature of the transition from pre-intensive acorn consumption, through high consumption, and finally to the arrival of Europeans and subsequent adoption of agriculture. Finally, through de novo metagenomic assembly of genomes, we uncovered hidden strain diversity among known and unknown oral taxa across pre-European contact Americans.

In chapter 4, “Reconstruction of ancient and modern bison oral microbiomes”, we present the first investigation of bison oral microbiomes from dental calculus, and the first modern non-human dental calculus to date. Unfortunately, host and oral microbial DNA was poorly preserved in Folsom era bison teeth, highlighting the necessity for further refinement of techniques for DNA extraction from herbivorous dental calculus of that age. However, modern bison calculus is enriched in *Actinomyces* and *Arachnia* oral microbes and has a high prevalence of the nitrate-reducing *Rothia aeria* and *Neisseria elongata*.

Bison calculus is also a rich source of host and environmental genetic material. Much of the herd suffered from parasitic infection, and this was detected in dental calculus metagenomes. However, with DNA read alignments confined to conserved regions, the detection of *Babesia* species could not be definitively established. We recovered DNA from the economically significant dietary species, *Sorghum bicolor*, and the plant fungal pathogen dispersed by bison, *Alternaria alternata*, with an even distribution of coverage. Bison are good candidates for innovating the reconstruction of ecological eukaryotic DNA from mammalian dental calculus.

Each chapter of this dissertation engages with the interconnectedness between oral microbial ecologies and their human and non-human hosts. With a nuanced approach to microbial functional dynamics, we can extrapolate patterns of microbial compositional shifts in response to host dietary practice and elucidate the larger, interconnected network encompassing human, animal, and environmental health. On a microscopic scale, we can observe host dynamics through time and across geographic expanses, in both humans and bison. Future innovation in extracting metagenomes from ancient herbivorous ruminant dental calculus could unlock a window into past ecological networks. Additionally, with the continued exploration of hidden microbial diversity in ancient oral microbiomes across the globe, we can expand our understanding of the impact of dietary transitions, colonialism, globalization, and modernization on oral microbiomes. Through these glimpses deep in time, molecular anthropologists can continue to explore the mechanisms mediating oral disease and elucidate the connections between host lifestyle and microbial composition.

BIBLIOGRAPHY

Abranches, J., Zeng, L., Kajfasz, J. K., Palmer, S. R., Chakraborty, B., Wen, Z. T., Richards, V. P., Brady, L. J., & Lemos, J. A. (2018). Biology of Oral Streptococci. *Microbiology Spectrum*, *6*(5), 10.1128/microbiolspec.gpp3-0042–2018. https://doi.org/10.1128/microbiolspec.gpp3-0042-2018

Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henrissat, B., White, O., Kelley, S. T., Methé, B., Schloss, P. D., Gevers, D., Mitreva, M., & Huttenhower, C. (2012). Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLoS Computational Biology*, *8*(6), e1002358. https://doi.org/10.1371/journal.pcbi.1002358

Abusleme, L., Dupuy, A. K., Dutzan, N., Silva, N., Burleson, J. A., Strausbaugh, L. D., Gamonal, J., & Diaz, P. I. (2013). The subgingival microbiome in health and periodontitis and its relationship with community biomass and inflammation. *The ISME Journal*, *7*(5), 1016–1025. https://doi.org/10.1038/ismej.2012.174

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J. A., Townsend, G., Sołtysiak, A., Alt, K. W., Parkhill, J., & Cooper, A. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics*, *45*(4), Article 4. https://doi.org/10.1038/ng.2536

Akersten, W. A., Foppe, T. M., & Jefferson, G. T. (1988). New Source of Dietary Data for Extinct Herbivores. *Quaternary Research*, *30*(1), 92–97. https://doi.org/10.1016/0033-5894(88)90090-7

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., Pollard, K. S., Sakharova, E., Parks, D. H., Hugenholtz, P., Segata, N., Kyrpides, N. C., & Finn, R. D. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology*, *39*(1), 105–114. https://doi.org/10.1038/s41587-020-0603-3

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, *62*(1), 245–253. https://doi.org/10.1111/j.1541-0420.2005.00440.x

Arellano, M. V., Leventhal, A., Guzman-Schmidt, S., Arellano Gomez, G. E., & Nijmeh, C. (2021). *An Ethnohistory of Santa Clara Valley and Adjacent Regions; Historic Ties of the Muwekma Ohlone Tribe of the San Francisco Bay Area and Tribal Stewardship Over the Human Remains Recovered on the Prometheus Project located at 575 Benton Street and Affiliated with the 3rd Mission Santa Clara de Thámien Indian Neophyte Cemetery and Indian Rancheria: Clareño Muwékma Ya Túnneste Nómmo [Where the Clareño Indians are Buried] Site CA-SCL-30/H*.

Armitage, P. L. (1975). The extraction and identification of opal phytoliths from the teeth of ungulates. *Journal of Archaeological Science*, *2*(3), 187–197. https://doi.org/10.1016/0305-4403(75)90056-4

Austin, R. M., Zuckerman, M., Honap, T. P., Lee, H., Ward, G. K., Warinner, C., Sankaranarayanan, K., & Hofman, C. A. (2022). Remembering St. Louis Individual—Structural violence and acute bacterial infections in a historical anatomical collection. *Communications Biology*, *5*(1), 1–10. https://doi.org/10.1038/s42003-022-03890-z

Bai, Y., Gangoiti, J., Dijkstra, B. W., Dijkhuizen, L., & Pijning, T. (2017). Crystal Structure of 4,6-α-Glucanotransferase Supports Diet-Driven Evolution of GH70 Enzymes from α-Amylases in Oral Bacteria. *Structure*, *25*(2), 231–242. https://doi.org/10.1016/j.str.2016.11.023

Bamforth, D. B. (1987). Historical Documents and Bison Ecology on the Great Plains. *Plains Anthropologist*, *32*(115), 1–16. https://doi.org/10.1080/2052546.1987.11909364

Bamforth, D. E. (1988). *Ecology and Human Organization on the Great Plains*. Plenum Press.

Bastias-Montes, J.-M., Flores-Varela, L.-E., Reyes-Calderón, O.-A., Vidal-San-Martín, C., Muñoz-Fariña, O., Quevedo-León, R., & Acuña-Nelson, S.-M. (2020). Teosinte (Dioon mejiae) Flour: Nutritional and Physicochemical Characterization of the Seed Flour of the Living Fossil in Honduras. *Agronomy*, *10*(4), Article 4. https://doi.org/10.3390/agronomy10040481

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*, *10*, e65088. https://doi.org/10.7554/eLife.65088

Belstrøm, D., Holmstrup, P., Nielsen, C. H., Kirkby, N., Twetman, S., Heitmann, B. L., Klepac-Ceraj, V., Paster, B. J., & Fiehn, N.-E. (2014). Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. *Journal of Oral Microbiology*, *6*, 10.3402/jom.v6.23609. https://doi.org/10.3402/jom.v6.23609

Bement, L. C. (1997). The Cooper Site: A Stratified Folsom Bison Kill in Oklahoma. *Plains Anthropologist*, *42*(159), 85–99. https://www.jstor.org/stable/25669445

Bement, L. C., & Carter, B. J. (2010). Jake Bluff: Clovis Bison Hunting on the Southern Plains of North America. *American Antiquity*, *75*(4), 907–933.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergmann, G. T., Craine, J. M., Ii, M. S. R., & Fierer, N. (2015). Seasonal Shifts in Diet and Gut Microbiota of the American Bison (Bison bison). *PLOS ONE*, *10*(11), e0142409. https://doi.org/10.1371/journal.pone.0142409

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E.,

Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., … Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, 1–12. https://doi.org/10.1038/s41587-023-01688-w

Bock, R., Jackson, L., de Vos, A., & Jorgensen, W. (2004). Babesiosis of cattle. *Parasitology*, *129 Suppl*, S247-269. https://doi.org/10.1017/s0031182004005190

Bolognini, D., Halgren, A., Lou, R. N., Raveane, A., Rocha, J. L., Guarracino, A., Soranzo, N., Chin, C.-S., Garrison, E., & Sudmant, P. H. (2024). Recurrent evolution and selection shape structural diversity at the amylase locus. *Nature*, *634*(8034), 617–625. https://doi.org/10.1038/s41586-024-07911-1

Borry, M., Cordova, B., Perri, A., Wibowo, M. C., Honap, T., Ko, W. T. J., Yu, J., Britton, K., Flink, L. G., Power, R. C., Stuijts, I., Garcia, D. S., Hofman, C. A., Hagan, R. W., Kagone, T. S., Meda, N., Carabin, H., Jacobson, D., Reinhard, K., … Warinner, C. (2019). CoproID predicts the source of coprolites and paleofeces using microbiome composition and host DNA content. *bioRxiv*, 871533. https://doi.org/10.1101/871533

Bos, K. I., Kühnert, D., Herbig, A., Esquivel-Gomez, L. R., Andrades Valtueña, A., Barquera, R., Giffin, K., Kumar Lankapalli, A., Nelson, E. A., Sabin, S., Spyrou, M. A., & Krause, J. (2019). Paleomicrobiology: Diagnosis and Evolution of Ancient Pathogens. *Annual Review of Microbiology*, *73*(1), 639–666. https://doi.org/10.1146/annurev-micro-090817-062436

Bosch, T. C. G., Wigley, M., Colomina, B., Bohannan, B., Meggers, F., Amato, K. R., Azad, M. B., Blaser, M. J., Brown, K., Dominguez-Bello, M. G., Ehrlich, S. D., Elinav, E., Finlay, B. B., Geddie, K., Geva-Zatorsky, N., Giles-Vernick, T., Gros, P., Guillemin, K., Haraoui, L.-P., … Melby, M. K. (2024). The potential importance of the built-environment microbiome and its impact on human health. *Proceedings of the National Academy of Sciences*, *121*(20), e2313971121. https://doi.org/10.1073/pnas.2313971121

Brealey, J. C., Leitão, H. G., Hofstede, T., Kalthoff, D. C., & Guschanski, K. (2021). The oral microbiota of wild bears in Sweden reflects the history of antibiotic use by humans. *Current Biology*, *31*(20), 4650-4658.e6. https://doi.org/10.1016/j.cub.2021.08.010

Brealey, J. C., Leitão, H. G., van der Valk, T., Xu, W., Bougiouri, K., Dalén, L., & Guschanski, K. (2020). Dental Calculus as a Tool to Study the Evolution of the Mammalian Oral Microbiome. *Molecular Biology and Evolution*, *37*(10), 3003–3022. https://doi.org/10.1093/molbev/msaa135

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, *38*(6), e87. https://doi.org/10.1093/nar/gkp1163

Brink, J. (2008). *Imagining Head-Smashed-In: Aboriginal Buffalo Hunting on the Northern Plains*. Athabasca University Press.

Brister, J. R., Ako-Adjei, D., Bao, Y., & Blinkova, O. (2015). NCBI viral genomes resource. *Nucleic Acids Research*, *43*(Database issue), D571-577. https://doi.org/10.1093/nar/gku1207

Bruserud, Ø., Siddiqui, H., Marthinussen, M. C., Chen, T., Jonsson, R., Oftedal, B. E., Olsen, I., Husebye, E. S., & Wolff, A. B. (2018). Oral microbiota in autoimmune polyendocrine syndrome type 1. *Journal of Oral Microbiology*, *10*(1), 1442986. https://doi.org/10.1080/20002297.2018.1442986

Buckley, A. A., Faustoferri, R. C., & Quivey, R. G. (2014). β-Phosphoglucomutase contributes to aciduricity in Streptococcus mutans. *Microbiology*, *160*(Pt 4), 818–827. https://doi.org/10.1099/mic.0.075754-0

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2017). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*, *9*(2), 410–419. https://doi.org/10.1111/2041-210X.12871

Casarin, R. C. V., Barbagallo, A., Meulman, T., Santos, V. R., Sallum, E. A., Nociti, F. H., Duarte, P. M., Casati, M. Z., & Gonçalves, R. B. (2013). Subgingival biodiversity in subjects with uncontrolled type-2 diabetes and chronic periodontitis. *Journal of Periodontal Research*, *48*(1), 30–36. https://doi.org/10.1111/j.1600-0765.2012.01498.x

CEC. (1997). *Ecological Regions of North America: Toward a common perspective*. Commission for Environmental Cooperation.

Chakraborty, B., Zeng, L., & Burne, R. A. (2022). The fruB Gene of Streptococcus mutans Encodes an Endo-Levanase That Enhances Growth on Levan and Influences Global Gene Expression. *Microbiology Spectrum*, *10*(3), e00522-22. https://doi.org/10.1128/spectrum.00522-22

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics*, *38*(23), 5315–5316. https://doi.org/10.1093/bioinformatics/btac672

Chen, J. Z., Kwong, Z., Gerardo, N. M., & Vega, N. M. (2024). Ecological drift during colonization drives within-host and between-host heterogeneity in an animal-associated symbiont. *PLOS Biology*, *22*(4), e3002304. https://doi.org/10.1371/journal.pbio.3002304

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, *2*(2), e107. https://doi.org/10.1002/imt2.107

Chhibber-Goel, J., Singhal, V., Bhowmik, D., Vivek, R., Parakh, N., Bhargava, B., & Sharma, A. (2016). Linkages between oral commensal bacteria and atherosclerotic plaques in coronary artery disease patients. *Npj Biofilms and Microbiomes*, *2*(1), 1–13. https://doi.org/10.1038/s41522-016-0009-7

Cohen, M. N., & Armelagos, G. (Eds.). (1984). *Paleopathology and the Origins of Agriculture*. Academic Press.

Cook, N. D. (1998). *Born to Die: Disease and New World Conquest, 1492-1650*. Cambridge University Press.

Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., O'Keefe, J. H., & Brand-Miller, J. (2005). Origins and evolution of the Western diet: Health implications for the 21st century. *The American Journal of Clinical Nutrition*, *81*(2), 341–354. https://doi.org/10.1093/ajcn.81.2.341

Corduneanu, A., Taulescu, M., Ursache, T. D., Ionică, A. M., & Mihalca, A. D. (2023). Piroplasms in farmed American bison, Bison bison from Romania. *Frontiers in Veterinary Science*, *10*. https://doi.org/10.3389/fvets.2023.1158072

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M., & Relman, D. A. (2012). The Application of Ecological Theory Toward an Understanding of the Human Microbiome. *Science*, *336*(6086), 1255–1262. https://doi.org/10.1126/science.1224203

Cotter, P. D., & Hill, C. (2003). Surviving the Acid Test: Responses of Gram-Positive Bacteria to Low pH. *Microbiology and Molecular Biology Reviews*, *67*(3), 429–453. https://doi.org/10.1128/mmbr.67.3.429-453.2003

Craine, J. M. (2021). Seasonal patterns of bison diet across climate gradients in North America. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-86260-9

Crosby, A. W. (1973). *The Columbian Exchange: Biological and Cultural Consequences of 1492*. Greenwood.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15758–15763. https://doi.org/10.1073/pnas.1314445110

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology*, *5*(7). https://doi.org/10.1101/cshperspect.a012567

Dahlquist-Axe, G., Standeven, F. J., Speller, C. F., Tedder, A., & Meehan, C. J. (2024). Inferring diet, disease and antibiotic resistance from ancient human oral microbiomes. *Microbial Genomics*, *10*(5). https://doi.org/10.1099/mgen.0.001251

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Davies, G., McCann, B., Sturdevant, J., Swenson, F., & Ovchinnikov, I. V. (2019). Isotopic paleoecology of Northern Great Plains bison during the Holocene. *Scientific Reports*, *9*, 16637. https://doi.org/10.1038/s41598-019-52873-4

de Oliveira, K., Asevedo, L., Calegari, M. R., Gelfo, J. N., Mothé, D., & Avilla, L. (2021). From oral pathology to feeding ecology: The first dental calculus paleodiet study of a South American native megamammal. *Journal of South American Earth Sciences*, *109*, 103281. https://doi.org/10.1016/j.jsames.2021.103281

DeMers, M. (2022). Alternaria alternata as endophyte and pathogen. *Microbiology*, *168*(3), 001153. https://doi.org/10.1099/mic.0.001153

DiGiuseppe, D., Grant, D., Mabie, E., Foo, H., & Leventhal, A. (2021). *Burial and Archaeological Data Recovery Program Conducted on a Portion of Thámien Rúmmeytak—Thámien (Guadalupe) River Site CA-SCL-128 Located in Downtown San Jose, Santa Clara County, California*. D&D Osteological Services, LLC.

Döbereiner, J., Dutra, I. S., Rosa, I. V., & Blobel, H. (2000). "Cara inchada" of cattle, an infectious, apparently soil antibiotics-dependant periodontitis in Brazil. *Pesquisa Veterinária Brasileira*, *20*, 47–64. https://doi.org/10.1590/S0100-736X2000000200001

Dominy, S. S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., Nguyen, M., Haditsch, U., Raha, D., Griffin, C., Holsinger, L. J., Arastu-Kapur, S., Kaba, S., Lee, A., Ryder, M. I., Potempa, B., Mydel, P., Hellvard, A., Adamowicz, K., … Potempa, J. (2019). Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Science Advances*, *5*(1), eaau3333. https://doi.org/10.1126/sciadv.aau3333

Douglas, K. C., Halbert, N. D., Kolenda, C., Childers, C., Hunter, D. L., & Derr, J. N. (2011). Complete mitochondrial DNA sequence analysis of Bison bison and bison-cattle hybrids: Function and phylogeny. *Mitochondrion*, *11*(1), 166–175. https://doi.org/10.1016/j.mito.2010.09.005

Driver, J. C., & Maxwell, D. (2013). Bison death assemblages and the interpretation of human hunting behaviour. *Quaternary International*, *297*, 100–109. https://doi.org/10.1016/j.quaint.2012.12.038

Eerkens, J. W., Nichols, R. V., Murray, G. G. R., Perez, K., Murga, E., Kaijankoski, P., Rosenthal, J. S., Engbring, L., & Shapiro, B. (2018). A probable prehistoric case of meningococcal disease from San Francisco Bay: Next generation sequencing of Neisseria meningitidis from dental calculus and osteological evidence. *International Journal of Paleopathology*, *22*, 173–180. https://doi.org/10.1016/j.ijpp.2018.05.001

Eerkens, J. W., Tichinin, A., Karapanos, N., Campbell-Grey, A., Firenzi, A. M., Bartelink, E. J., Leventhal, A. M., Arellano, M. V., & DeOrnellas, B. (2022). *Contrasting Male and Female Dietary Life Histories: A Case Study at an Ancestral Muwekma Ohlone Heritage Site in San Jose, California*.

Eisenhofer, R., Kanzawa-Kiriyama, H., Shinoda, K., & Weyrich, L. S. (2020). Investigating the demographic history of Japan using ancient oral microbiota. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1812), 20190578. https://doi.org/10.1098/rstb.2019.0578

Esberg, A., Barone, A., Eriksson, L., Lif Holgerson, P., Teneberg, S., & Johansson, I. (2020). Corynebacterium matruchotii Demography and Adhesion Determinants in the Oral Cavity of Healthy Individuals. *Microorganisms*, *8*(11), 1780. https://doi.org/10.3390/microorganisms8111780

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., & Lemon, K. P. (2018). New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): A Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems*, *3*(6), 10.1128/msystems.00187-18. https://doi.org/10.1128/msystems.00187-18

Fan, X., Alekseyenko, A. V., Wu, J., Peters, B. A., Jacobs, E. J., Gapstur, S. M., Purdue, M. P., Abnet, C. C., Stolzenberg-Solomon, R., Miller, G., Ravel, J., Hayes, R. B., & Ahn, J. (2018). Human oral microbiome and prospective risk for pancreatic cancer: A population-based nested case-control study. *Gut*, *67*(1), 120–127. https://doi.org/10.1136/gutjnl-2016-312580

Fellows Yates, J. A., Andrades Valtueña, A., Vågene, Å. J., Cribdon, B., Velsko, I. M., Borry, M., Bravo-Lopez, M. J., Fernandez-Guerra, A., Green, E. J., Ramachandran, S. L., Heintzman, P. D., Spyrou, M. A., Hübner, A., Gancz, A. S., Hider, J., Allshouse, A. F., Zaro, V., & Warinner, C. (2021). Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir. *Scientific Data*, *8*(1), 31. https://doi.org/10.1038/s41597-021-00816-y

Fellows Yates, J. A., Velsko, I. M., Aron, F., Posth, C., Hofman, C. A., Austin, R. M., Parker, C. E., Mann, A. E., Nägele, K., Arthur, K. W., Arthur, J. W., Bauer, C. C., Crevecoeur, I., Cupillard, C., Curtis, M. C., Dalén, L., Díaz-Zorita Bonilla, M., Díez Fernández-Lomana, J. C., Drucker, D. G., … Warinner, C. (2021). The evolution and changing ecology of the African hominid oral microbiome. *Proceedings of the National Academy of Sciences*, *118*(20), e2021655118. https://doi.org/10.1073/pnas.2021655118

Fierer, N., Ladau, J., Clemente, J. C., Leff, J. W., Owens, S. M., Pollard, K. S., Knight, R., Gilbert, J. A., & McCulley, R. L. (2013). Reconstructing the Microbial Diversity and Function of Pre-Agricultural Tallgrass Prairie Soils in the United States. *Science*, *342*(6158), 621–624. https://doi.org/10.1126/science.1243768

Figueroa-Gonzalez, P. A., Bornemann, T. L. V., Adam, P. S., Plewka, J., Révész, F., von Hagen, C. A., Táncsics, A., & Probst, A. J. (2020). Saccharibacteria as Organic Carbon Sinks in Hydrocarbon-Fueled Communities. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.587782

Fleskes, R. E., Johnson, S. J., Honap, T. P., Abin, C. A., Gilmore, J. K., Oubré, L., Bueschgen, W. D., Abel, S. M., Ofunniyin, A. A., Lewis, C. M., & Schurr, T. G. (2024). Oral microbial diversity in 18th century African individuals from South Carolina. *Communications Biology*, *7*(1), 1213. https://doi.org/10.1038/s42003-024-06893-0

Formosinho, J., Bencard, A., & Whiteley, L. (2022). Environmentality in biomedicine: Microbiome research and the perspectival body. *Studies in History and Philosophy of Science*, *91*, 148–158. https://doi.org/10.1016/j.shpsa.2021.11.005

Foster, J. W. (2004). Escherichia coli acid resistance: Tales of an amateur acidophile. *Nature Reviews Microbiology*, *2*(11), 898–907. https://doi.org/10.1038/nrmicro1021

Gancz, A. S., Farrer, A. G., Nixon, M. P., Wright, S., Arriola, L., Adler, C., Davenport, E. R., Gully, N., Cooper, A., Britton, K., Dobney, K., Silverman, J. D., & Weyrich, L. S. (2023). Ancient dental calculus reveals oral microbiome shifts associated with lifestyle and disease in Great Britain. *Nature Microbiology*, *8*(12), Article 12. https://doi.org/10.1038/s41564-023-01527-3

Gardner, K. S., Bartelink, E. J., Martinez, A., Leventhal, A., & Cambra, R. (2018). Breastfeeding and weaning practices of the ancestral Ohlone Indians of California: A case study using stable isotope analysis of bone collagen. *International Journal of Osteoarchaeology*, *28*(5), 523–534. https://doi.org/10.1002/oa.2681

Garland, C. J., Turner, B. L., & Klaus, H. D. (2016). Biocultural Consequences of Spanish Contact in the Lambayeque Valley Region of Northern Peru: Internal Enamel Micro-Defects as Indicators of Early Life Stress. *International Journal of Osteoarchaeology*, *26*(6), 947–958. https://doi.org/10.1002/oa.2505

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gluckman, P. D., Low, F. M., Buklijas, T., Hanson, M. A., & Beedle, A. S. (2011). How evolutionary principles improve the understanding of human health and disease. *Evolutionary Applications*, *4*(2), 249–263. https://doi.org/10.1111/j.1752-4571.2010.00164.x

Granehäll, L., Huang, K. D., Tett, A., Manghi, P., Paladin, A., O'Sullivan, N., Rota-Stabelli, O., Segata, N., Zink, A., & Maixner, F. (2021). Metagenomic analysis of ancient dental calculus reveals unexplored diversity of oral archaeal Methanobrevibacter. *Microbiome*, *9*(1), 197. https://doi.org/10.1186/s40168-021-01132-8

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Grzeczka, A., Lech, M., Wozniak, G., Graczyk, S., Kordowitzki, P., Olejnik, M., Gehrke, M., & Jaśkowski, J. M. (2023). Periodontitis Disease in Farmed Ruminants—Current State of Research. *International Journal of Molecular Sciences*, *24*(11), 9763. https://doi.org/10.3390/ijms24119763

Guan, N., & Liu, L. (2020). Microbial response to acid stress: Mechanisms and applications. *Applied Microbiology and Biotechnology*, *104*(1), 51–65. https://doi.org/10.1007/s00253-019-10226-1

Hajishengallis, G., Darveau, R. P., & Curtis, M. A. (2012). The Keystone Pathogen Hypothesis. *Nature Reviews. Microbiology*, *10*(10), 717–725. https://doi.org/10.1038/nrmicro2873

Halvorsrud, K., Lewney, J., Craig, D., & Moynihan, P. J. (2019). Effects of Starch on Oral Health: Systematic Review to Inform WHO Guideline. *Journal of Dental Research*, *98*(1), 46–53. https://doi.org/10.1177/0022034518788283

Hill, J. D. (1998). Violent Encounters: Ethnogenesis and Ethnocide in Long-Term Contact Situations. In *Studies in Culture Contact: Interaction, Culture Change, and Archaeology*. Southern Illinois University Press.

Hillman, J. D., Andrews, S. W., & Dzuback, A. L. (1987). Acetoin production by wild-type strains and a lactate dehydrogenase-deficient mutant of Streptococcus mutans. *Infection and Immunity*, *55*(6), 1399. https://doi.org/10.1128/iai.55.6.1399-1402.1987

HMP. (2012). A framework for human microbiome research. *Nature*, *486*(7402), 215–221. https://doi.org/10.1038/nature11209

Honap, T. P., Monroe, C. R., Johnson, S. J., Jacobson, D. K., Abin, C. A., Austin, R. M., Sandberg, P., Levine, M., Sankaranarayanan, K., & Lewis Jr., C. M. (2023). Oral

metagenomes from Native American Ancestors reveal distinct microbial lineages in the pre-contact era. *American Journal of Biological Anthropology*, *n/a*(n/a). https://doi.org/10.1002/ajpa.24735

Huang, S., Li, J. W., Zheng, L. W., Qiao, W. W., & McGrath, C. (2024). One Health and Oral Health: A Scoping Review to Inform Research and Present Challenges. *JDR Clinical & Translational Research*, *9*(1_suppl), 88S-98S. https://doi.org/10.1177/23800844241273821

Innocenti, G., Martino, M. E., Stellini, E., Di Fiore, A., & Quagliariello, A. (2023). Dental calculus microbiome correlates with dietary intake. *Molecular Oral Microbiology*, *38*(3), 189–197. https://doi.org/10.1111/omi.12404

Isenberg, A. C. (2001). *The Destruction of the Bison: An Environmental History, 1750-1920*. Cambridge University Press.

Iyer, R., Iverson, T. M., Accardi, A., & Miller, C. (2002). A biological role for prokaryotic ClC chloride channels. *Nature*, *419*(6908), 715–718. https://doi.org/10.1038/nature01000

Jackson, I., Woodman, P., Dowd, M., Fibiger, L., & Cassidy, L. M. (2024). Ancient Genomes From Bronze Age Remains Reveal Deep Diversity and Recent Adaptive Episodes for Human Oral Pathobionts. *Molecular Biology and Evolution*, *41*(3), msae017. https://doi.org/10.1093/molbev/msae017

Jacobson, D. K., Honap, T. P., Monroe, C., Lund, J., Houk, B. A., Novotny, A. C., Robin, C., Marini, E., & Lewis, C. M. (2020). Functional diversity of microbial ecologies estimated from ancient human coprolites and dental calculus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1812), 20190586. https://doi.org/10.1098/rstb.2019.0586

Jansen, M. G. S., Borsanelli, A. C., Dutra, I. S., & Ubiali, D. G. (2022). Pathology of chronic ovine periodontitis. *Pesquisa Veterinária Brasileira*, *42*, e07170. https://doi.org/10.1590/1678-5150-PVB-7170

Jensen, E. D., Selway, C. A., Allen, G., Bednarz, J., Weyrich, L. S., Gue, S., Peña, A. S., & Couper, J. (2021). Early markers of periodontal disease and altered oral microbiota are associated with glycemic control in children with type 1 diabetes. *Pediatric Diabetes*, *22*(3), 474–481. https://doi.org/10.1111/pedi.13170

Jersie-Christensen, R. R., Lanigan, L. T., Lyon, D., Mackie, M., Belstrøm, D., Kelstrup, C. D., Fotakis, A. K., Willerslev, E., Lynnerup, N., Jensen, L. J., Cappellini, E., & Olsen, J. V. (2018). Quantitative metaproteomics of medieval dental calculus reveals individual oral health status. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-07148-3

Jin, Y., & Yip, H.-K. (2002). Supragingival calculus: Formation and control. *Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists*, *13*(5), 426–441. https://doi.org/10.1177/154411130201300506

Johnson, E., & Bement, L. C. (2009). Bison butchery at Cooper, a Folsom site on the Southern Plains. *Journal of Archaeological Science*, *36*(7), 1430–1446. https://doi.org/10.1016/j.jas.2009.02.007

Johnston, E. R., Rodriguez-R, L. M., Luo, C., Yuan, M. M., Wu, L., He, Z., Schuur, E. A. G., Luo, Y., Tiedje, J. M., Zhou, J., & Konstantinidis, K. T. (2016). Metagenomics Reveals

Pervasive Bacterial Populations and Reduced Community Diversity across the Alaska Tundra Ecosystem. *Frontiers in Microbiology*, *7*, 579. https://doi.org/10.3389/fmicb.2016.00579

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics (Oxford, England)*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Jorns, T., Craine, J., Towne, E. G., & Knox, M. (2020). Climate structures bison dietary quality and composition at the continental scale. *Environmental DNA*, *2*(1), e47. https://doi.org/10.1002/edn3.47

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population scale DNA sequence data. *Genome Research*, gr.176552.114. https://doi.org/10.1101/gr.176552.114

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, *42*(Database issue), D199-205. https://doi.org/10.1093/nar/gkt1076

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457-462. https://doi.org/10.1093/nar/gkv1070

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*, e7359. https://doi.org/10.7717/peerj.7359

Kato, I., Vasquez, A. A., Moyerbrailean, G., Land, S., Sun, J., Lin, H.-S., & Ram, J. L. (2016). Oral microbiome and history of smoking and colorectal cancer. *Journal of Epidemiological Research*, *2*(2), Article 2. https://doi.org/10.5430/jer.v2n2p92

Kato, I., Vasquez, A., Moyerbrailean, G., Land, S., Djuric, Z., Sun, J., Lin, H.-S., & Ram, J. L. (2017). Nutritional Correlates of Human Oral Microbiome. *Journal of the American College of Nutrition*, *36*(2), 88–98. https://doi.org/10.1080/07315724.2016.1185386

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC135756/

Kazarina, A., Petersone-Gordina, E., Kimsis, J., Kuzmicka, J., Zayakin, P., Grišķjans, Ž., Gerhards, G., & Ranka, R. (2021). The Postmedieval Latvian Oral Microbiome in the Context of Modern Dental Calculus and Modern Dental Plaque Microbial Profiles. *Genes*, *12*(2), Article 2. https://doi.org/10.3390/genes12020309

Kennett, D. J., Lipson, M., Prufer, K. M., Mora-Marín, D., George, R. J., Rohland, N., Robinson, M., Trask, W. R., Edgar, H. H. J., Hill, E. C., Ray, E. E., Lynch, P., Moes, E., O'Donnell, L., Harper, T. K., Kate, E. J., Ramos, J., Morris, J., Gutierrez, S. M., … Reich, D. (2022). South-to-north migration preceded the advent of intensive farming

in the Maya region. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-29158-y

Kennett, D. J., Prufer, K. M., Culleton, B. J., George, R. J., Robinson, M., Trask, W. R., Buckley, G. M., Moes, E., Kate, E. J., Harper, T. K., O'Donnell, L., Ray, E. E., Hill, E. C., Alsgaard, A., Merriman, C., Meredith, C., Edgar, H. J. H., Awe, J. J., & Gutierrez, S. M. (2020). Early isotopic evidence for maize as a staple grain in the Americas. *Science Advances*, *6*(23). https://doi.org/10.1126/sciadv.aba3245

Kianoush, N., Adler, C. J., Nguyen, K.-A. T., Browne, G. V., Simonian, M., & Hunter, N. (2014). Bacterial Profile of Dentine Caries and the Impact of pH on Bacterial Population Diversity. *PLOS ONE*, *9*(3), e92940. https://doi.org/10.1371/journal.pone.0092940

Klapper, M., Hübner, A., Ibrahim, A., Wasmuth, I., Borry, M., Haensch, V. G., Zhang, S., Al-Jammal, W. K., Suma, H., Fellows Yates, J. A., Frangenberg, J., Velsko, I. M., Chowdhury, S., Herbst, R., Bratovanov, E. V., Dahse, H.-M., Horch, T., Hertweck, C., González Morales, M. R., … Stallforth, P. (2023). Natural products from reconstructed bacterial genomes of the Middle and Upper Paleolithic. *Science*, *380*(6645), 619–624. https://doi.org/10.1126/science.adf5300

Klaus, H. D., & Alvarez-Calderón, R. (2017). Escaping Conquest? A First Look at Regional Cultural and Biological Variation in Postcontact Eten, Peru. In M. S. Murphy & H. D. Klaus (Eds.), *Colonized Bodies, Worlds Transformed: Toward a Global Bioarchaeology of Contact and Colonialism* (p. 0). University Press of Florida. https://doi.org/10.5744/florida/9780813060750.003.0004

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., Bushman, F. D., Knight, R., & Kelley, S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, *8*(9), 761–763. https://doi.org/10.1038/nmeth.1650

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., & Ding, L. (2009). VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, *25*(17), 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

Kolodny, O., & Schulenburg, H. (2020). Microbiome-mediated plasticity directs host evolution along several distinct time scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1808), 20190589. https://doi.org/10.1098/rstb.2019.0589

Krulwich, T. A., Sachs, G., & Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nature Reviews. Microbiology*, *9*(5), 330–343. https://doi.org/10.1038/nrmicro2549

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10.1101/gr.092759.109

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Lahti, L., & Shetty, S. (2012). *Microbiome R package* [Dataset]. http://microbiome.github.io

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology*, *30*, 191–202. https://doi.org/10.1007/s10682-016-9821-z

Lamont, R. J., Koo, H., & Hajishengallis, G. (2018). The oral microbiota: Dynamic communities and host interactions. *Nature Reviews. Microbiology*, *16*(12), 745–759. https://doi.org/10.1038/s41579-018-0089-x

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), Article 4. https://doi.org/10.1038/nmeth.1923

Lassalle, F., Spagnoletti, M., Fumagalli, M., Shaw, L., Dyble, M., Walker, C., Thomas, M. G., Bamberg Migliano, A., & Balloux, F. (2018). Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet. *Molecular Ecology*, *27*(1), 182–195. https://doi.org/10.1111/mec.14435

Leatherman, T., & Goodman, A. (2020). Building on the biocultural syntheses: 20 years and still expanding. *American Journal of Human Biology*, *32*(4), e23360. https://doi.org/10.1002/ajhb.23360

Leventhal, A., Atwood, M., Grant, D., Cambra, R., Nijmeh, C., Arellano, M. V., Guzman-Schmidt, S., Gomez, G. E., & Sanchez, N. (2011). *Final Report on the Burial and Archaeological Data Recovery Program Conducted on a Portion of the Mission Santa Clara Indian Neophyte Cemetery (1781-1818): Clareño Muwékma Ya Túnnešte Nómmo [Where the Clareño Indians are Buried] Site (CA-SCL-30/H), Located in the City of Santa Clara, Santa Clara County, California*. Muwekma Ohlone Tribe/Ohlone Families Consulting Services.

Leventhal, A., Atwood, M., Grant, D., Morley, S., Cambra, R., Field, L., Nijmeh, C., Arellano, M. V., Rodriguez, S., Guzman-Schmidt, S., Gomez, G. E., & Sanchez, N. (2010). *Final Report on the Burial and Arcaheological Data Recovery Program Conducted on a Portion of a Middle Period Ohlone Indian Cemetery, Yuki Kutsuimi Šaatoš Inūxw [Sand Hill Road] Sites: CA-SCL-287 and CA-SMA-263*. Stanford University.

Leventhal, A., DiGiuseppe, D., Grant, D., Cambra, R., Arellano, M. V., Guzman-Schmidt, S., Gomez, G. E., & Sanchez, A. (2017). *Report on the Analysis and Temporal Placement of an Ancestral Muwekma Ohlone Burial Recovered from 'Ayttakiš 'Éete Hiramwiš Trépam-tak [Place of Woman Sleeping Under the Pipe Site], CA-ALA-677/H Located in the Town of Sunol, Alameda County, California*. San Francisco Public Utilities Commission.

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, *73*(6), 1943–1967. https://doi.org/10.2307/1941447

Lewontin, R. C. (1983). Gene, organism and environment. In D. S. Bendall (Ed.), *Evolution from molecules to men*. Cambridge University Press.

Li, B.-Z., Zhou, H.-Y., Guo, B., Chen, W.-J., Tao, J.-H., Cao, N.-W., Chu, X.-J., & Meng, X. (2020). Dysbiosis of oral microbiota is associated with systemic lupus erythematosus. *Archives of Oral Biology*, *113*, 104708. https://doi.org/10.1016/j.archoralbio.2020.104708

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn

graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033
Li, H. (2011). *Wgsim-Read simulator for next generation sequencing* [Computer software].
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, J. (2020). *Metagenomic Sequencing of California Dental Calculus Samples from CA-SCL-134 and Yuki Kutsuimi Saatos Inuxw (CA-SCL-287)* [Honors College Undergraduate Research Thesis]. University of Oklahoma.
Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. *2*.
Lieverse, A. R. (1999). Diet and the aetiology of dental calculus. *International Journal of Osteoarchaeology*, *9*(4), 219–232. https://doi.org/10.1002/(SICI)1099-1212(199907/08)9:4<219::AID-OA475>3.0.CO;2-V
Lightfoot, K. G., Cuthrell, R. Q., Striplen, C. J., & Hylkema, M. G. (2013). Rethinking the Study of Landscape Management Practices Among Hunter-Gatherers in North America. *American Antiquity*, *78*(2), 285–301.
Lock, M. (2017). Recovering the Body. *Annual Review of Anthropology*, *46*(Volume 46, 2017), 1–14. https://doi.org/10.1146/annurev-anthro-102116-041253
Loesche, W. J. (1986). Role of Streptococcus mutans in human dental decay. *Microbiological Reviews*, *50*(4), 353–380. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC373078/
Lommi, S., Manzoor, M., Engberg, E., Agrawal, N., Lakka, T. A., Leinonen, J., Kolho, K.-L., & Viljakainen, H. (2022). The Composition and Functional Capacities of Saliva Microbiota Differ Between Children With Low and High Sweet Treat Consumption. *Frontiers in Nutrition*, *9*, 864687. https://doi.org/10.3389/fnut.2022.864687
Lott, D. F. (2002). *American Bison: A Natural History* (1st ed.). University of California Press. https://www.jstor.org/stable/10.1525/j.ctt1ppxd1
Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science*, *3*, e104. https://doi.org/10.7717/peerj-cs.104
Lund, P. A., De Biase, D., Liran, O., Scheler, O., Mira, N. P., Cetecioglu, Z., Fernández, E. N., Bover-Cid, S., Hall, R., Sauer, M., & O'Byrne, C. (2020). Understanding How Microorganisms Respond to Acid pH Is Central to Their Control and Successful Exploitation. *Frontiers in Microbiology*, *11*, 556140. https://doi.org/10.3389/fmicb.2020.556140
Lund, P., Tramonti, A., & De Biase, D. (2014). Coping with low pH: Molecular strategies in neutralophilic bacteria. *FEMS Microbiology Reviews*, *38*(6), 1091–1125. https://doi.org/10.1111/1574-6976.12076
Ma, Z., Richard, H., & Foster, J. W. (2003). pH-Dependent Modulation of Cyclic AMP Levels and GadW-Dependent Repression of RpoS Affect Synthesis of the GadX Regulator and Escherichia coli Acid Resistance. *Journal of Bacteriology*, *185*(23), 6852–6859. https://doi.org/10.1128/jb.185.23.6852-6859.2003
Maixner, F., Turaev, D., Cazenave-Gassiot, A., Janko, M., Krause-Kyora, B., Hoopmann, M. R., Kusebauch, U., Sartain, M., Guerriero, G., O'Sullivan, N., Teasdale, M., Cipollini,

G., Paladin, A., Mattiangeli, V., Samadelli, M., Tecchiati, U., Putzer, A., Palazoglu, M., Meissen, J., … Zink, A. (2018). The Iceman's Last Meal Consisted of Fat, Wild Meat, and Cereals. *Current Biology*, *28*(14), 2348-2355.e9. https://doi.org/10.1016/j.cub.2018.05.067

Mann, A. E., Fellows Yates, J. A., Fagernäs, Z., Austin, R. M., Nelson, E. A., & Hofman, C. A. (2023). Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus. *Quaternary International*, *653–654*, 33–46. https://doi.org/10.1016/j.quaint.2020.11.019

Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Fellows Yates, J. A., Salazar-García, D. C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G. R., Stone, A. C., Lewis, C. M., Krause, J., Hofman, C., … Warinner, C. (2018). Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-28091-9

Marsh, P. D. (1994). Microbial ecology of dental plaque and its significance in health and disease. *Advances in Dental Research*, *8*(2), 263–271. https://doi.org/10.1177/08959374940080022001

Mazurel, D., Carda-Diéguez, M., Langenburg, T., Žiemytė, M., Johnston, W., Martínez, C. P., Albalat, F., Llena, C., Al-Hebshi, N., Culshaw, S., Mira, A., & Rosier, B. T. (2023). Nitrate and a nitrate-reducing Rothia aeria strain as potential prebiotic or synbiotic treatments for periodontitis. *Npj Biofilms and Microbiomes*, *9*(1), 1–12. https://doi.org/10.1038/s41522-023-00406-3

McMillan, N. A., Fuhlendorf, S. D., Luttbeg, B., Goodman, L. E., Davis, C. A., Allred, B. W., & Hamilton, R. G. (2022). Bison movements change with weather: Implications for their continued conservation in the Anthropocene. *Ecology and Evolution*, *12*(12), e9586. https://doi.org/10.1002/ece3.9586

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Merchant, S., Wood, D. E., & Salzberg, S. L. (2014). Unexpected cross-species contamination in genome sequencing projects. *PeerJ*, *2*, e675. https://doi.org/10.7717/peerj.675

Miklossy, J. (2016). Bacterial Amyloid and DNA are Important Constituents of Senile Plaques: Further Evidence of the Spirochetal and Biofilm Nature of Senile Plaques. *Journal of Alzheimer's Disease*, *53*(4), 1459–1473. https://doi.org/10.3233/JAD-160451

Milani, C., Alessandri, G., Mancabelli, L., Mangifesta, M., Lugli, G. A., Viappiani, A., Longhi, G., Anzalone, R., Duranti, S., Turroni, F., Ossiprandi, M. C., van Sinderen, D., & Ventura, M. (2020). Multi-omics Approaches To Decipher the Impact of Diet and Host Physiology on the Mammalian Gut Microbiome. *Applied and Environmental Microbiology*, *86*(23), e01864-20. https://doi.org/10.1128/AEM.01864-20

Millen, A. E., Dahhan, R., Freudenheim, J. L., Hovey, K. M., Li, L., McSkimming, D. I., Andrews, C. A., Buck, M. J., LaMonte, M. J., Kirkwood, K. L., Sun, Y., Murugaiyan, V.,

Tsompana, M., & Wactawski-Wende, J. (2022). Dietary carbohydrate intake is associated with the subgingival plaque oral microbiome abundance and diversity in a cohort of postmenopausal women. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-06421-2

Milliken, R., Fitzgerald, R., Hylkema, M., Groza, R., Origer, T., Bieling, D., Leventhal, A., Wiberg, R., Gottsfield, A., Gillette, D., Bellifemine, V., Strother, E., Cartier, R., & A.Fredrickson, D. (2007). Punctuated Culture Change in the San Francisco Bay Area. In *California Prehistory: Colonization, Culture and Complexity* (pp. 99–125).

Modi, A., Attolini, D., Zaro, V., Pisaneschi, L., Innocenti, G., Vai, S., Caramelli, D., Moggi Cecchi, J., Quagliariello, A., Mariotti Lippi, M., & Lari, M. (2023). Combined metagenomic and archaeobotanical analyses on human dental calculus: A cross-section of lifestyle conditions in a Copper Age population of central Italy. *Quaternary International*, *653–654*, 69–81. https://doi.org/10.1016/j.quaint.2021.12.003

Modi, A., Pisaneschi, L., Zaro, V., Vai, S., Vergata, C., Casalone, E., Caramelli, D., Moggi-Cecchi, J., Mariotti Lippi, M., & Lari, M. (2020). Combined methodologies for gaining much information from ancient dental calculus: Testing experimental strategies for simultaneously analysing DNA and food residues. *Archaeological and Anthropological Sciences*, *12*(1), 10. https://doi.org/10.1007/s12520-019-00983-5

Monroe, C. (2014). *Correlating Biological Relationships, Social Inequality, and Population Movement among Prehistoric California Foragers: Ancient Human DNA Analysis from CA-SCL-38 (Yukisma Site)*. UC Santa Barbara.

Moraitou, M., Forsythe, A., Fellows Yates, J. A., Brealey, J. C., Warinner, C., & Guschanski, K. (2022). Ecology, Not Host Phylogeny, Shapes the Oral Microbiome in Closely Related Species. *Molecular Biology and Evolution*, *39*(12), msac263. https://doi.org/10.1093/molbev/msac263

Mummert, A., Esche, E., Robinson, J., & Armelagos, G. J. (2011). Stature and robusticity during the agricultural transition: Evidence from the bioarchaeological record. *Economics & Human Biology*, *9*(3), 284–301. https://doi.org/10.1016/j.ehb.2011.03.004

Murphy, M. S., & Klaus, H. D. (2017). *Colonized Bodies, Worlds Transformed: Toward A Global Bioarchaeology of Contact and Colonialism*. University Press of Florida. https://doi.org/10.2307/j.ctvx0725r

Nasidze, I., Li, J., Schroeder, R., Creasey, J. L., Li, M., & Stoneking, M. (2011). High Diversity of the Saliva Microbiome in Batwa Pygmies. *PLoS ONE*, *6*(8), e23352. https://doi.org/10.1371/journal.pone.0023352

National Institutes of Health. (2021). *Oral Health in America: Advances and Challenges*. National Institute of Dental and Craniofacial Research.

Naud, S., Valles, C., Abdillah, A., Abou Chacra, L., Mekhalif, F. Z., Ibrahim, A., Caputo, A., Baudoin, J.-P., Gouriet, F., Bittar, F., Lagier, J.-C., Ranque, S., Fenollar, F., Tidjani Alou, M., & Raoult, D. (2023). Preliminary landscape of Candidatus Saccharibacteria in the human microbiome. *Frontiers in Cellular and Infection Microbiology*, *13*. https://doi.org/10.3389/fcimb.2023.1195679

Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A. G., Chatelier, E. L., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., Santos, M. B. Q. dos, Blom, N., Borruel, N., … Ehrlich, S. D. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, *32*(8), 822–828. https://doi.org/10.1038/nbt.2939

Niven, L. B., & Glenn Hill, M. (1998). Season of Bison Mortality at Three Plains Archaic Kill Sites in Wyoming. *Plins Anthropologist*, *163*, 5–26.

Nuss, E. T., & Tanumihardjo, S. A. (2010). Maize: A Paramount Staple Crop in the Context of Global Nutrition. *Comprehensive Reviews in Food Science and Food Safety*, *9*(4), 417–436. https://doi.org/10.1111/j.1541-4337.2010.00117.x

Obregon-Tito, A. J., Tito, R. Y., Metcalf, J., Sankaranarayanan, K., Clemente, J. C., Ursell, L. K., Zech Xu, Z., Van Treuren, W., Knight, R., Gaffney, P. M., Spicer, P., Lawson, P., Marin-Reyes, L., Trujillo-Villarroel, O., Foster, M., Guija-Poma, E., Troncoso-Corzo, L., Warinner, C., Ozga, A. T., & Lewis, C. M. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Communications*, *6*, 6505. https://doi.org/10.1038/ncomms7505

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (1996). Niche Construction. *The American Naturalist*, *147*(4), 641–648.

Oh, J., Byrd, A. L., Deming, C., Conlan, S., Kong, H. H., & Segre, J. A. (2014). Biogeography and individuality shape function in the human skin metagenome. *Nature*, *514*(7520), 59–64. https://doi.org/10.1038/nature13786

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *32*(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2001). *vegan: Community Ecology Package* (p. 2.6-8) [Dataset]. https://doi.org/10.32614/CRAN.package.vegan

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733-745. https://doi.org/10.1093/nar/gkv1189

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), 2864–2868. https://doi.org/10.1038/ismej.2017.126

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1), 132. https://doi.org/10.1186/s13059-016-0997-x

Oppenheimer, J., Rosen, B. D., Heaton, M. P., Vander Ley, B. L., Shafer, W. R., Schuetze, F. T., Stroud, B., Kuehn, L. A., McClure, J. C., Barfield, J. P., Blackburn, H. D., Kalbfleisch, T. S., Bickhart, D. M., Davenport, K. M., Kuhn, K. L., Green, R. E., Shapiro, B., & Smith, T. P. L. (2021). A Reference Genome Assembly of American Bison, Bison bison bison. *Journal of Heredity*, *112*(2), 174–183. https://doi.org/10.1093/jhered/esab003

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-020-00011-0

Orosz, F. (2015). Two recently sequenced vertebrate genomes are contaminated with apicomplexan species of the Sarcocystidae family. *International Journal for Parasitology*, *45*(13), 871–878. https://doi.org/10.1016/j.ijpara.2015.07.002

Ottoni, C., Borić, D., Cheronet, O., Sparacello, V., Dori, I., Coppa, A., Antonović, D., Vujević, D., Price, T. D., Pinhasi, R., & Cristiani, E. (2021). Tracking the transition to agriculture in Southern Europe through ancient DNA analysis of dental calculus. *Proceedings of the National Academy of Sciences*, *118*(32). https://doi.org/10.1073/pnas.2102116118

Ottoni, C., Guellil, M., Ozga, A. T., Stone, A. C., Kersten, O., Bramanti, B., Porcier, S., & Van Neer, W. (2019). Metagenomic analysis of dental calculus in ancient Egyptian baboons. *Scientific Reports*, *9*, 19637. https://doi.org/10.1038/s41598-019-56074-x

Ovchinnikov, I. V., & McCann, B. (2023). Mitogenomes revealed the history of bison colonization of Northern Plains after the Last Glacial Maximum. *Scientific Reports*, *13*(1), 11417. https://doi.org/10.1038/s41598-023-37599-8

Ozga, A. T., Gilby, I., Nockerts, R. S., Wilson, M. L., Pusey, A., & Stone, A. C. (2019). Oral microbiome diversity in chimpanzees from Gombe National Park. *Scientific Reports*, *9*(1), 17354. https://doi.org/10.1038/s41598-019-53802-1

Ozga, A. T., Nieves-Colón, M. A., Honap, T. P., Sankaranarayanan, K., Hofman, C. A., Milner, G. R., Lewis Jr., C. M., Stone, A. C., & Warinner, C. (2016). Successful enrichment and recovery of whole mitochondrial genomes from ancient human dental calculus. *American Journal of Physical Anthropology*, *160*(2), 220–228. https://doi.org/10.1002/ajpa.22960

Ozga, A. T., & Ottoni, C. (2023). Dental calculus as a proxy for animal microbiomes. *Quaternary International*, *653–654*, 47–52. https://doi.org/10.1016/j.quaint.2021.06.012

Pääbo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(6), 1939–1943. https://doi.org/10.1073/pnas.86.6.1939

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Panich, L., Armstrong, B., Galvan, A., & Canzonieri, C. (n.d.). *Identifying Ohlone Ancestors Buried at San Pedro y San Pablo: Insights from Mission Records and Archaeology*.

Panich, L. M., Arellano, M. V., Wilcox, M., Flores, G., & Connell, S. (2024). Fighting erasure and dispossession in the San Francisco Bay Area: Putting archaeology to work for the Muwekma Ohlone Tribe. *Frontiers in Environmental Archaeology*, *3*. https://doi.org/10.3389/fearc.2024.1394106

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2022). GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1), D785–D794. https://doi.org/10.1093/nar/gkab776

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., & Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649-662.e20. https://doi.org/10.1016/j.cell.2019.01.001

Peck, T. R. (2004). *Bison Ethology and Native Settlement Patterns during the Old Women's Phase on the Northwestern Plains*. British Archaeological Reports.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: Efficient ancient genome reconstruction. *Genome Biology*, *17*(1), 60. https://doi.org/10.1186/s13059-016-0918-z

Perez-Mon, C., Qi, W., Vikram, S., Frossard, A., Makhalanyane, T., Cowan, D., & Frey, B. 2021. (2021). Shotgun metagenomics reveals distinct functional diversity and metabolic capabilities between 12 000-year-old permafrost and active layers on Muot da Barba Peider (Swiss Alps). *Microbial Genomics*, *7*(4), 000558. https://doi.org/10.1099/mgen.0.000558

Plumb, G. E., White, P. J., Coughenour, M. B., & Wallen, R. L. (2009). Carrying capacity, migration, and dispersal in Yellowstone bison. *Biological Conservation*, *142*(11), 2377–2387. https://doi.org/10.1016/j.biocon.2009.05.019

Poco, S. E., Nakazawa, F., Sato, M., & Hoshino, E. (1996). Eubacterium minutum sp. Nov., isolated from human periodontal pockets. *International Journal of Systematic Bacteriology*, *46*(1), 31–34. https://doi.org/10.1099/00207713-46-1-31

Polley, H. W., & Collins, S. L. (1984). Relationships of Vegetation and Environment in Buffalo Wallows. *The American Midland Naturalist*, *112*(1), 178–186. https://doi.org/10.2307/2425471

Prufer, K. M., Robinson, M., & Kennett, D. J. (2021). Terminal Pleistocene Through Middle Holocene Occupations in Southeastern Mesoamerica: Linking Ecology and Culture in he Context of Neotropical Foragers and Early Farmers. *Ancient Mesoamerica*, *32*, 439–460.

Putrino, A., Marinelli, E., Galeotti, A., Ferrazzano, G. F., Ciribè, M., & Zaami, S. (2024). A Journey into the Evolution of Human Host-Oral Microbiome Relationship through Ancient Dental Calculus: A Scoping Review. *Microorganisms*, *12*(5), 902. https://doi.org/10.3390/microorganisms12050902

Quagliariello, A., Modi, A., Innocenti, G., Zaro, V., Conati Barbaro, C., Ronchitelli, A., Boschin, F., Cavazzuti, C., Dellù, E., Radina, F., Sperduti, A., Bondioli, L., Ricci, S., Lognoli, M., Belcastro, M. G., Mariotti, V., Caramelli, D., Mariotti Lippi, M., Cristiani, E., … Lari, M. (2022). Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-34416-0

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., & Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience*, *8*(9), giz107. https://doi.org/10.1093/gigascience/giz107

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology*, *62*(2), 142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Rampelli, S., Schnorr, S. L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., Brigidi, P., Crittenden, A. N., Henry, A. G., & Candela, M. (2015). Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Current Biology: CB*, *25*(13), 1682–1693. https://doi.org/10.1016/j.cub.2015.04.055

Ratajczak, Z., Collins, S. L., Blair, J. M., Koerner, S. E., Louthan, A. M., Smith, M. D., Taylor, J. H., & Nippert, J. B. (2022). Reintroducing bison results in long-running and resilient increases in grassland diversity. *Proceedings of the National Academy of Sciences*, *119*(36), e2210433119. https://doi.org/10.1073/pnas.2210433119

Ray, E. E., Neff, N. C., Lynch, P., Mes, J., Lachniet, M. S., Kennett, D. J., & Prufer, K. M. (2024). The development of early farming diets and population change in the Maya region and their climate context. *Quaternary International*, *689–690*, 66–78. https://doi.org/10.1016/j.quaint.2023.09.008

Rodriguez, M., & Makalowski, W. (2022). Mobilome of Apicomplexa Parasites. *Genes*, *13*(5), 887. https://doi.org/10.3390/genes13050887

Rohart, F., Gautier, B., Singh, A., & Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660). https://doi.org/10.1098/rstb.2013.0624

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Saleh, A. A., Kraisitudomsook, N., Frank, E. E., Dendy, S. P., Leslie, J. F., & Garrett, K. A. (2023). Phytobiome Stampede: Bison as Potential Dispersal Agents for the Tallgrass

Prairie Microbiome. *PhytoFrontiers™*, *3*(3), 512–517. https://doi.org/10.1094/PHYTOFR-01-23-0004-SC

Santamaria, R. M., Estrada, K., López, M. E., Rojas, E., Martínez, G., Alcalá, Y., Rojas, C., Álvarez, J. A., Lira, J. J., Santamaria, T. V., Sánchez-Flores, A., & Figueroa, J. V. (2024). Comparative Transcriptome Analysis of Babesia bigemina Attenuated Vaccine and Virulent Strains of Mexican Origin. *Vaccines*, *12*(3), Article 3. https://doi.org/10.3390/vaccines12030309

Sarkar, A., McInroy, C. J. A., Harty, S., Raulo, A., Ibata, N. G. O., Valles-Colomer, M., Johnson, K. V.-A., Brito, I. L., Henrich, J., Archie, E. A., Barreiro, L. B., Gazzaniga, F. S., Finlay, B. B., Koonin, E. V., Carmody, R. N., & Moeller, A. H. (2024). Microbial transmission in the social microbiome and host health and disease. *Cell*, *187*(1), 17–43. https://doi.org/10.1016/j.cell.2023.12.014

Sawafuji, R., Saso, A., Suda, W., Hattori, M., & Ueda, S. (2020). Ancient DNA analysis of food remains in human dental calculus from the Edo period, Japan. *PLOS ONE*, *15*(3), e0226654. https://doi.org/10.1371/journal.pone.0226654

Scardina, G. A., & Messina, P. (2012). Good Oral Health and Diet. *BioMed Research International*, *2012*(1), 720692. https://doi.org/10.1155/2012/720692

Scholz, C. F. P., & Kilian, M. (2016). The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus Propionibacterium to the proposed novel genera Acidipropionibacterium gen. Nov., Cutibacterium gen. Nov. And Pseudopropionibacterium gen. Nov. *International Journal of Systematic and Evolutionary Microbiology*, *66*(11), 4422–4432. https://doi.org/10.1099/ijsem.0.001367

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*(1), 88. https://doi.org/10.1186/s13104-016-1900-2

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T. S., Shapiro, N., Blood, P. D., Gurevich, A., Bai, Y., Turaev, D., … McHardy, A. C. (2017). Critical Assessment of Metagenome Interpretation—A benchmark of metagenomics software. *Nature Methods*, *14*(11), 1063–1071. https://doi.org/10.1038/nmeth.4458

Sedghi, L., DiMassa, V., Harrington, A., Lynch, S. V., & Kapila, Y. L. (2021). *The oral microbiome: Role of key organisms and complex networks in oral health and disease—Sedghi—2021—Periodontology 2000—Wiley Online Library*. https://onlinelibrary.wiley.com/doi/full/10.1111/prd.12393

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Sen, T., Cawthon, C. R., Ihde, B. T., Hajnal, A., DiLorenzo, P. M., de La Serre, C. B., & Czaja, K. (2017). Diet-driven microbiota dysbiosis is associated with vagal remodeling and obesity. *Physiology & Behavior*, *173*, 305–317. https://doi.org/10.1016/j.physbeh.2017.02.027

Shaiber, A., Willis, A. D., Delmont, T. O., Roux, S., Chen, L.-X., Schmid, A. C., Yousef, M., Watson, A. R., Lolans, K., Esen, Ö. C., Lee, S. T. M., Downey, N., Morrison, H. G.,

Dewhirst, F. E., Mark Welch, J. L., & Eren, A. M. (2020). Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *Genome Biology*, *21*(1), 292. https://doi.org/10.1186/s13059-020-02195-w

Shaw, J., & Yu, Y. W. (2023). Fast and robust metagenomic sequence comparison through sparse chaining with skani. *Nature Methods*, *20*(11), 1661–1665. https://doi.org/10.1038/s41592-023-02018-3

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Shiferaw, B., Prasanna, B. M., Hellin, J., & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, *3*(3), 307–327. https://doi.org/10.1007/s12571-011-0140-5

Silbergeld, E. K. (2019). One health and the agricultural transition in food animal production. *Global Transitions*, *1*, 83–92. https://doi.org/10.1016/j.glt.2019.01.003

Simpson, B., & Hilken, T. (2024). *Conservation planning with bison producers. Grazing Technical Note No. 190-BIO-94*. U.S. Department of Agriculture, Natural Resources Conservation Service. https://directives.nrcs.usda.gov/sites/default/files2/1732536012/TN-190-BIO-94%20Conservation%20Planning%20with%20Bison%20Producers.pdf

Singer, M. (1996). Farewell to Adaptationism: Unnatural Selection and the Politics of Biology. *Medical Anthropology Quarterly*, *10*(4), 496–515. https://www.jstor.org/stable/648657

Sipes, K., Almatari, A., Eddie, A., Williams, D., Spirina, E., Rivkina, E., Liang, R., Onstott, T. C., Vishnivetskaya, T. A., & Lloyd, K. G. (2021). Eight Metagenome-Assembled Genomes Provide Evidence for Microbial Adaptation in 20,000- to 1,000,000-Year-Old Siberian Permafrost. *Applied and Environmental Microbiology*, *87*(19), e00972-21. https://doi.org/10.1128/AEM.00972-21

Siqueira, J. F., & Rôças, I. N. (2003). Pseudoramibacter alactolyticus in Primary Endodontic Infections. *Journal of Endodontics*, *29*(11), 735–738. https://doi.org/10.1097/00004770-200311000-00012

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences*, *111*(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Socransky, S. S., & Haffajee, A. D. (2005). Periodontal microbial ecology. *Periodontology 2000*, *38*(1), 135–187. https://doi.org/10.1111/j.1600-0757.2005.00107.x

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., & Kent, R. L. (1998). Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*, *25*(2), 134–144. https://doi.org/10.1111/j.1600-051x.1998.tb02419.x

Sonnenburg, E. D., & Sonnenburg, J. L. (2014). Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. *Cell Metabolism*, *20*(5), 779–786. https://doi.org/10.1016/j.cmet.2014.07.003

Speth, J. D. (2020). *Paleoindian Bison Hunting on the North American Great Plains—Two Critical Nutritional Constraints*.

Standeven, F. J., Dahlquist-Axe, G., Speller, C. F., Meehan, C. J., & Tedder, A. (2024). *An efficient pipeline for creating metagenomic-assembled genomes from ancient oral microbiomes*. https://doi.org/10.1101/2024.09.18.613623

Steffens, T. S., & Finnis, E. (2022). Context matters: Leveraging anthropology within one health. *One Health*, *14*, 100393. https://doi.org/10.1016/j.onehlt.2022.100393

Steinegger, M., & Salzberg, S. L. (2020). Terminating contamination: Large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology*, *21*(1), 115. https://doi.org/10.1186/s13059-020-02023-1

Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., & Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature Biotechnology*, *37*(8), Article 8. https://doi.org/10.1038/s41587-019-0202-3

Stroupe, S., Forgacs, D., Harris, A., Derr, J. N., & Davis, B. W. (2022). Genomic evaluation of hybridization in historic and modern North American Bison (Bison bison). *Scientific Reports*, *12*(1), 6397. https://doi.org/10.1038/s41598-022-09828-z

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., & Wu, C. H. (2015). UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, *31*(6), 926–932. https://doi.org/10.1093/bioinformatics/btu739

Taschereau Mamers, D. (2019). Human-Bison Relations As Sites of Settler Colonial Violence and Decolonial Resurgence. *Humanimalia*, *10*(2), 10–41. https://doi.org/10.52537/humanimalia.9500

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, *44*(14), 6614–6624. https://doi.org/10.1093/nar/gkw569

Teles, R., & Wang, C.-Y. (2011). Mechanisms involved in the association between peridontal diseases and cardiovascular disease. *Oral Diseases*, *17*(5), 450–461. https://doi.org/10.1111/j.1601-0825.2010.01784.x

Thomas, M., Webb, M., Ghimire, S., Blair, A., Olson, K., Fenske, G. J., Fonder, A. T., Christopher-Hennings, J., Brake, D., & Scaria, J. (2017). Metagenomic characterization of the effect of feed additives on the gut microbiome and antibiotic resistome of feedlot cattle. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-12481-6

Todd, L. C., Hofman, J. L., & Schultz, C. B. (1990). Seasonality of the Scottsbluff and Lipscomb Bison Bonebeds: Implications for Modeling Paleoindian Subsistence. *American Antiquity*, *55*(4), 813–827. https://doi.org/10.2307/281252

Tong, F., Wang, T., Gao, N. L., Liu, Z., Cui, K., Duan, Y., Wu, S., Luo, Y., Li, Z., Yang, C., Xu, Y., Lin, B., Yang, L., Pauciullo, A., Shi, D., Hua, G., Chen, W.-H., & Liu, Q. (2022). The microbiome of the buffalo digestive tract. *Nature Communications*, *13*(1), 823. https://doi.org/10.1038/s41467-022-28402-9

Vartoukian, S. R., Adamowska, A., Lawlor, M., Moazzez, R., Dewhirst, F. E., & Wade, W. G. (2016). In Vitro Cultivation of 'Unculturable' Oral Bacteria, Facilitated by Community Culture and Media Supplementation with Siderophores. *PLoS ONE*, *11*(1), e0146926. https://doi.org/10.1371/journal.pone.0146926

Velsko, I. M., Fagernäs, Z., Tromp, M., Bedford, S., Buckley, H. R., Clark, G., Dudgeon, J., Flexner, J., Galipaud, J.-C., Kinaston, R., Lewis, C. M., Matisoo-Smith, E., Nägele, K., Ozga, A. T., Posth, C., Rohrlach, A. B., Shing, R., Simanjuntak, T., Spriggs, M., … Warinner, C. (2024). Exploring the potential of dental calculus to shed light on past human migrations in Oceania. *Nature Communications*, *15*(1), 10191. https://doi.org/10.1038/s41467-024-53920-z

Velsko, I. M., Fellows Yates, J. A., Aron, F., Hagan, R. W., Frantz, L. A. F., Loe, L., Martinez, J. B. R., Chaves, E., Gosden, C., Larson, G., & Warinner, C. (2019). Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome*, *7*(1), 102. https://doi.org/10.1186/s40168-019-0717-3

Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G., & Warinner, C. (2018). Selection of Appropriate Metagenome Taxonomic Classifiers for Ancient Microbiome Research. *mSystems*, *3*(4). https://doi.org/10.1128/mSystems.00080-18

Velsko, I. M., Semerau, L., Inskip, S. A., García-Collado, M. I., Ziesemer, K., Ruber, M. S., Benítez de Lugo Enrich, L., Molero García, J. M., Valle, D. G., Peña Ruiz, A. C., Salazar-García, D. C., Hoogland, M. L. P., & Warinner, C. (2022). Ancient dental calculus preserves signatures of biofilm succession and interindividual variation independent of dental pathology. *PNAS Nexus*, *1*(4), pgac148. https://doi.org/10.1093/pnasnexus/pgac148

Velsko, I. M., & Warinner, C. (2025). Streptococcus abundance and oral site tropism in humans and non-human primates reflects host and lifestyle differences. *Npj Biofilms and Microbiomes*, *11*(1), 1–13. https://doi.org/10.1038/s41522-024-00642-1

Walde, D. (2006). Bison breeding characteristics and interpretation of archaeological seasonality revisited. *International Journal of Osteoarchaeology*, *16*(6), 481–492. https://doi.org/10.1002/oa.852

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344. https://doi.org/10.1038/ng.2906

Warinner, C., Speller, C., & Collins, M. J. (2015). A new era in palaeomicrobiology: Prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1660), 20130376. https://doi.org/10.1098/rstb.2013.0376

Wendt, J. A. F., McWethy, D. B., Widga, C., & Shuman, B. N. (2022). Large-scale climatic drivers of bison distribution and abundance in North America since the Last Glacial Maximum. *Quaternary Science Reviews*, *284*, 107472. https://doi.org/10.1016/j.quascirev.2022.107472

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V., Farrell, M., Farrer, A. G., Francken, M., Gully, N., Haak, W., Hardy, K., Harvati, K., Held, P., Holmes, E. C., … Cooper, A. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, *544*(7650), Article 7650. https://doi.org/10.1038/nature21674

Whiley, R. A., & Hardie, J. M. (2015). Streptococcus. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–86). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118960608.gbm00612

White, D. J. (1997). Dental calculus: Recent insights into occurrence, formation, prevention, removal and oral health effects of supragingival and subgingival deposits. *European Journal of Oral Sciences*, *105*(5), 508–522. https://doi.org/10.1111/j.1600-0722.1997.tb00238.x

Wibowo, M. C., Yang, Z., Borry, M., Hübner, A., Huang, K. D., Tierney, B. T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., Martínez-Hernández, A., Luber, J. M., Kirstahler, P., Blohm, T., Smiley, F. E., Arnold, R., Ballal, S. A., Pamp, S. J., Russ, J., … Kostic, A. D. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature*, *594*(7862), Article 7862. https://doi.org/10.1038/s41586-021-03532-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Willems, A., & Collins, M. D. (2015). Pseudoramibacter. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–8). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118960608.gbm00631

Willis, J. R., & Gabaldón, T. (2020). The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems. *Microorganisms*, *8*(2), Article 2. https://doi.org/10.3390/microorganisms8020308

Wohlgemuth, E. (1996). Resource Intensification in Prehistoric Central California: Evidence from Archaeobotanical Data. *Journal of California and Great Basin Anthropology*, *18*(1), 81–103. https://www.jstor.org/stable/27825599

Wolf, M. (2015). Is there really such a thing as "one health"? Thinking about a more than human world from the perspective of cultural anthropology. *Social Science & Medicine*, *129*, 5–11. https://doi.org/10.1016/j.socscimed.2014.06.018

Woo, P. C. Y., Tam, D. M. W., Leung, K.-W., Lau, S. K. P., Teng, J. L. L., Wong, M. K. M., & Yuen, K.-Y. (2002). Streptococcus sinensis sp. Nov., a Novel Species Isolated from a Patient with Infective Endocarditis. *Journal of Clinical Microbiology*, *40*(3), 805. https://doi.org/10.1128/JCM.40.3.805-810.2002

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

World Health Organization. (2022). *Global oral health status report: Towards universal health coverage for oral health by 2030*.

Yamagishi, J., Asada, M., Hakimi, H., Tanaka, T. Q., Sugimoto, C., & Kawazu, S. (2017). Whole-genome assembly of Babesia ovata and comparative genomics between closely related pathogens. *BMC Genomics*, *18*(1), 832. https://doi.org/10.1186/s12864-017-4230-4

Yang, N., Wang, Y., Liu, X., Jin, M., Vallebueno-Estrada, M., Calfee, E., Chen, L., Dilkes, B. P., Gui, S., Fan, X., Harper, T. K., Kennett, D. J., Li, W., Lu, Y., Ding, J., Chen, Z., Luo, J., Mambakkam, S., Menon, M., … Ross-Ibarra, J. (2023). Two teosintes made modern maize. *Science*, *382*(6674), eadg8940. https://doi.org/10.1126/science.adg8940

Zaugg, J. L., & Kuttler, K. L. (1987). Experimental infections of Babesia bigemina in American bison. *Journal of Wildlife Diseases*, *23*(1), 99–102. https://doi.org/10.7589/0090-3558-23.1.99

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W., Tang, S., Xu, X., … Wang, J. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Medicine*, *21*(8), 895–905. https://doi.org/10.1038/nm.3914

Ziesemer, K. A., Mann, A. E., Sankaranarayanan, K., Schroeder, H., Ozga, A. T., Brandt, B. W., Zaura, E., Waters-Rist, A., Hoogland, M., Salazar-García, D. C., Aldenderfer, M., Speller, C., Hendy, J., Weston, D. A., MacDonald, S. J., Thomas, G. H., Collins, M. J., Lewis, C. M., Hofman, C., & Warinner, C. (2015). Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports*, *5*, 16498. https://doi.org/10.1038/srep16498

SUPPLEMENTARY MATERIAL A

Author List and Affiliations

Sarah J. Johnson[1,2], Tanvi P. Honap[1,2], Paige A. Lynch[3], Emily Moes[3,4], Erin Ray[3,7] José Mes[5], Douglas J. Kennett[6], Keith M. Prufer[3, 7], and Cecil M. Lewis, Jr. [1, 2]

[1] Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, Norman, OK 73019, U.S.A.

[2] Department of Anthropology, University of Oklahoma, Norman, OK 73069, U.S.A.

[3] Department of Anthropology, University of New Mexico, Albuquerque, NM 87131, U.S.A.

[4] Physician Assistant Studies, University of St. Frances, Albuquerque, NM 87107, U.S.A.

[5] Uchbenkah K'in Ajaw Association, Santa Cruz, Belize

[6] Department of Anthropology, University of California Santa Barbara

[7] Center for Stable Isotopes, University of New Mexico

Author Contributions

Conceptualization: C.M.L, T.P.H., K.M.P.

Resources: K.M.P., D.J.K, J.M.

Data curation: S.J.J., P.A.L., E.M., E.R.

Formal analysis: S.J.J., T.P.H.

Funding acquisition: C.M.L., T.P.H.

Investigation: T.P.H., S.J.J.

Methodology: S.J.J., T.P.H.

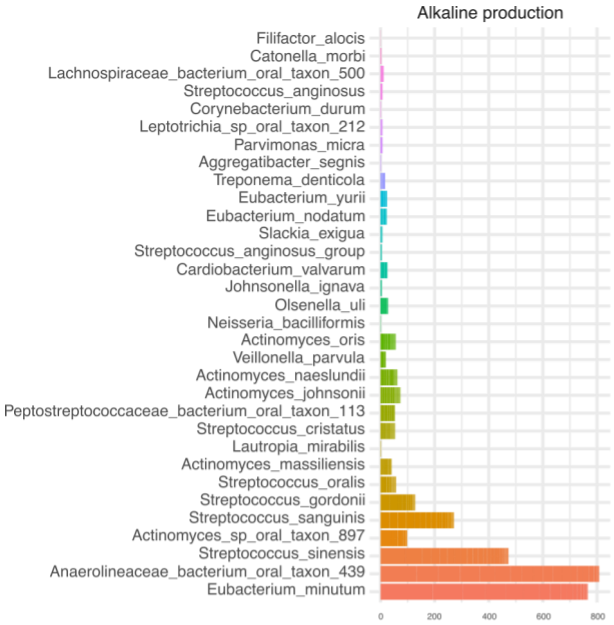
Supervision: C.M.L, T.P.H

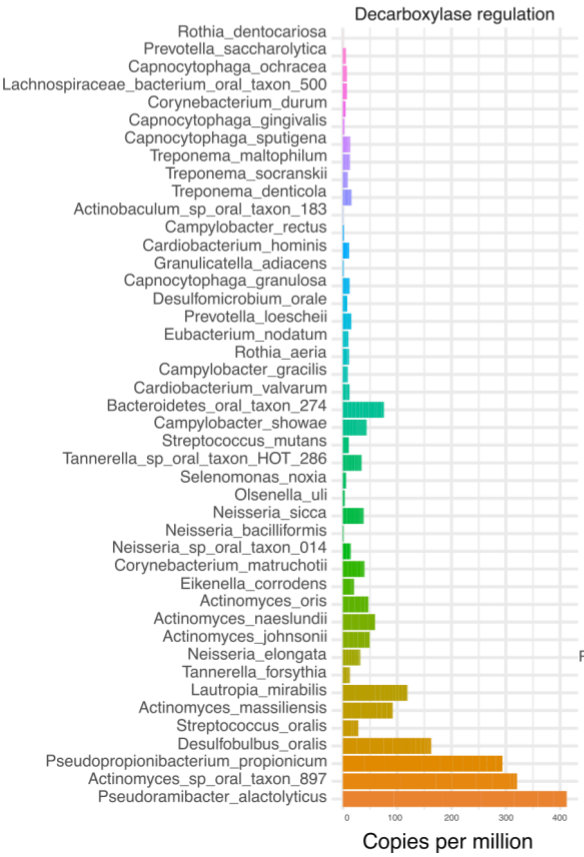
Visualization: S.J.J.

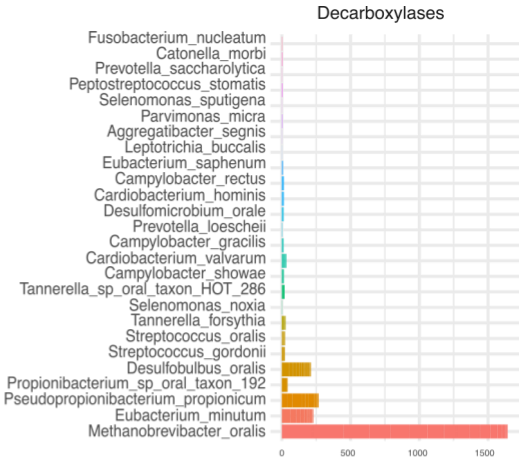
Writing – original draft: S.J.J.

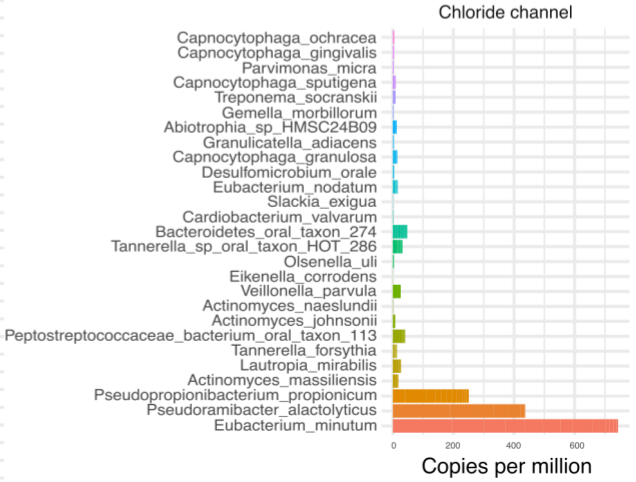
Writing – review and editing: All authors

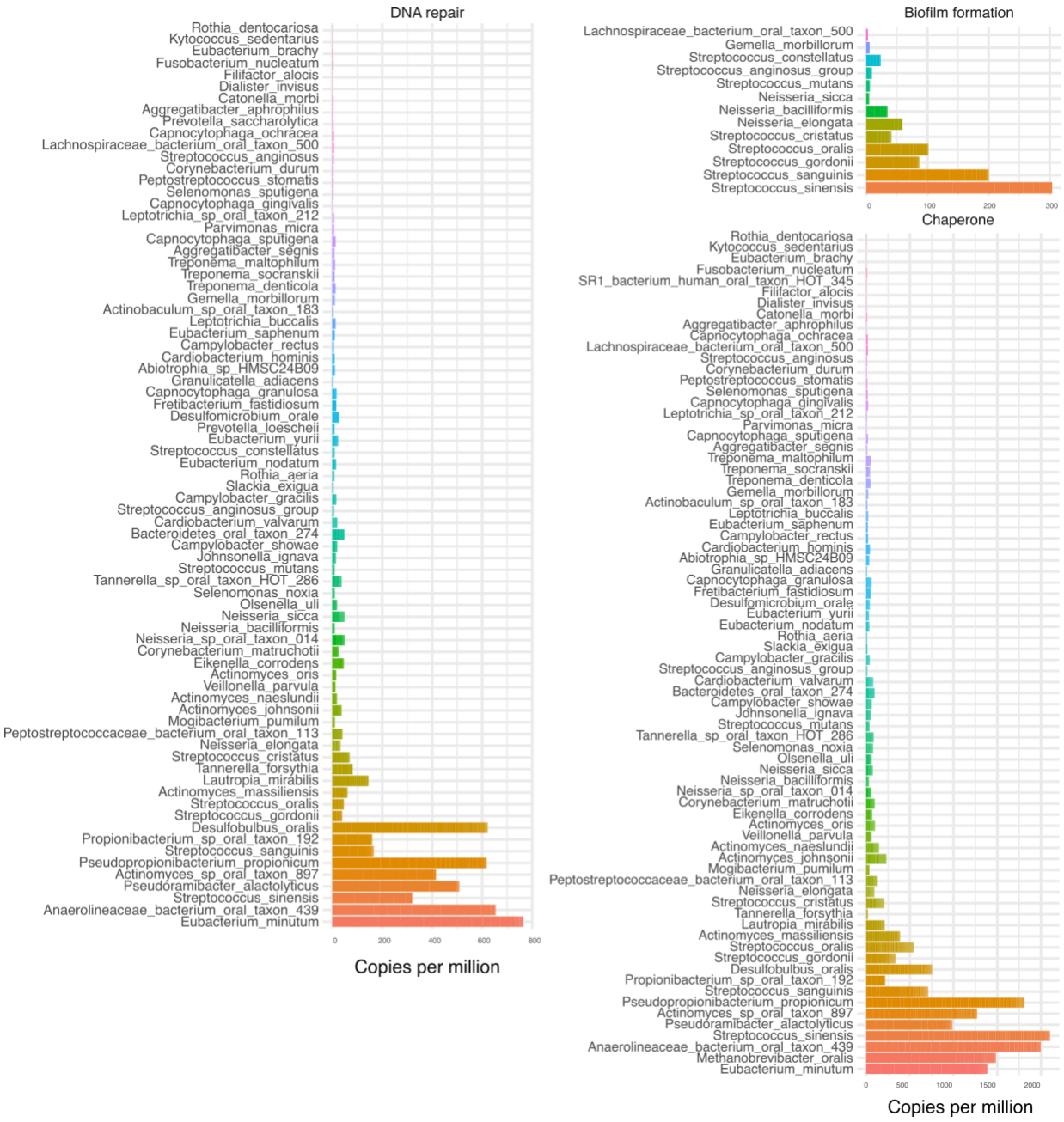
Supplementary Figures







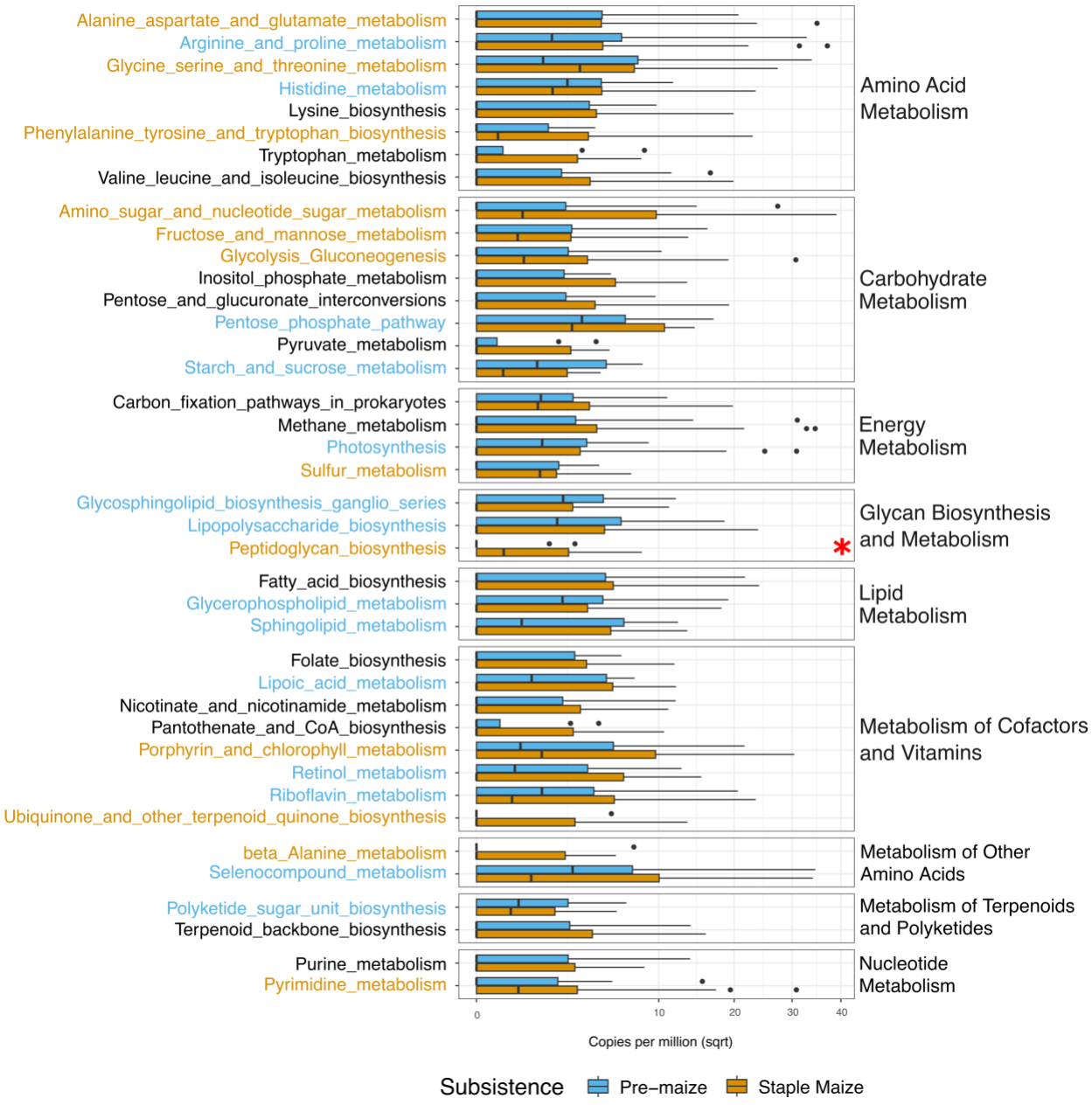




Supplementary Figure 1 Distribution of ATR mechanisms across taxa.

ATR mechanisms identified through the literature. Quantified from annotated UniRef50 database.

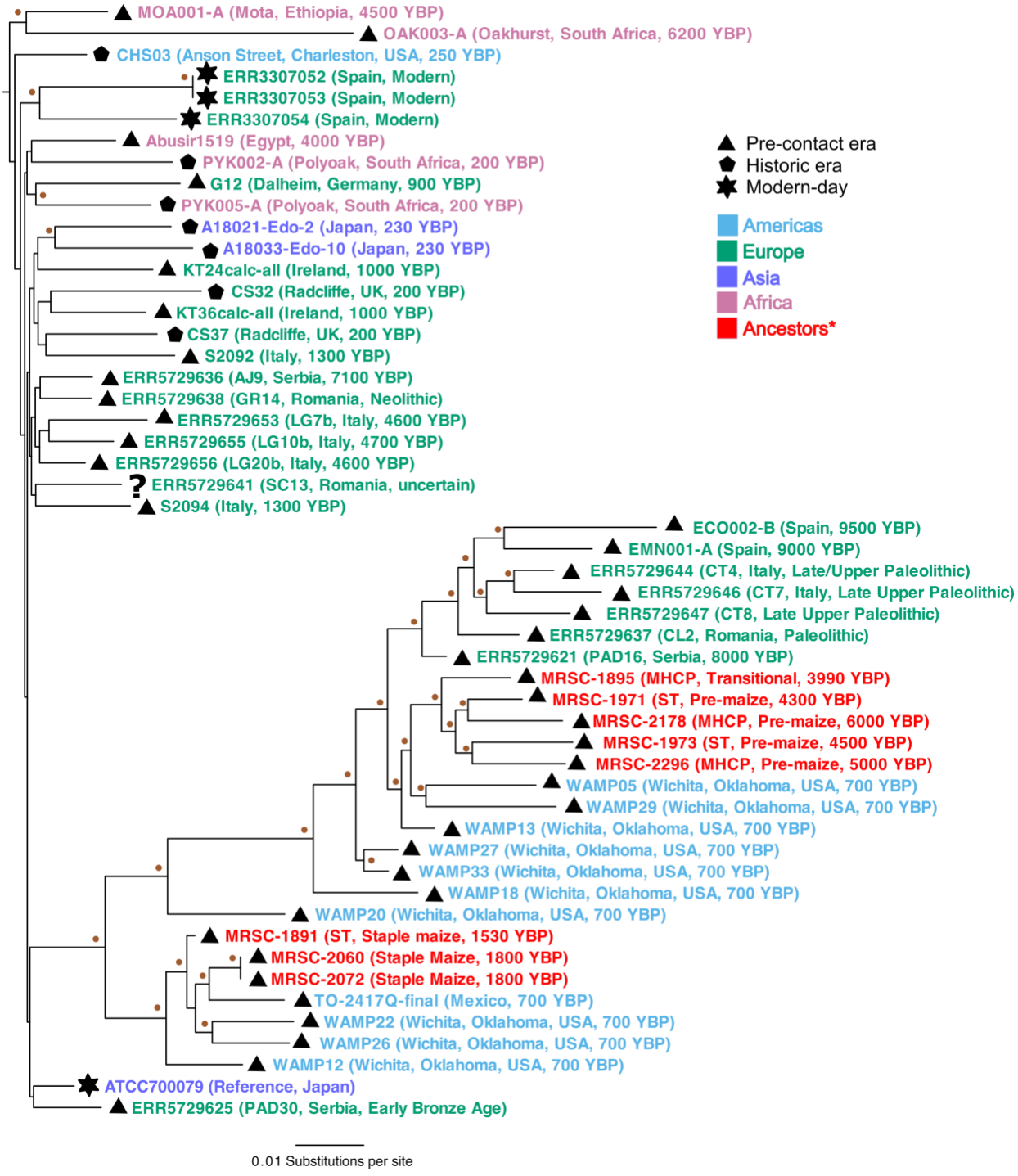
Differential Tannerella forsythia functions



Supplementary Figure 2 Differential abundance of T. forsythia metabolic functions.

Differential abundance of *T. forsythia* metabolic functions. Significance of peptidoglycan biosynthesis (red asterisk) calculated by Pairwise Wilcoxon test (* $p < 0.05$). Metabolic functions colored by subsistence pattern in which it is elevated.

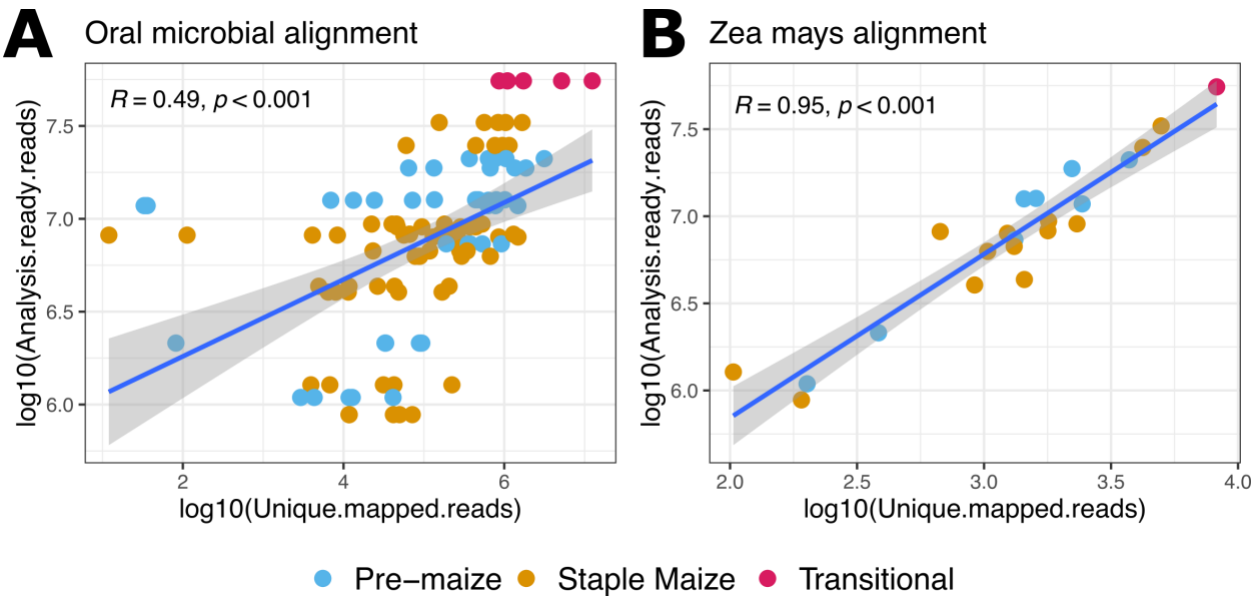
Eubacterium minutum



Supplementary Figure 3 Maximum likelihood tree for *Eubacterium minutum*.

The tree is based on an alignment of 22,656 homozygous bi-allelic SNPs, allowing up to 10% of missing data. The tree was built using the TVM+F+ASC+R2 model in IQ-TREE and rooted using a strain recovered from a wild-born chimpanzee. Brown circles indicate

greater than 95% bootstrap support estimated using 1,000 ultrafast bootstrap replicates. Sample names are colored by geographic origin and accompanied by a symbol indicating time period.



Supplementary Figure 4 Correlation between AR reads and uniquely mapped reads.

Differences in correlation between the number of uniquely mapping reads to A) oral microbes and B) *Z. mays*. Pearson correlation (*R*) indicated on each plot.

Supplementary Tables

Context	Subsistence
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Pre-agricultural	Pre-maize
Transitional	Transitional
Pre-agricultural	Pre-maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Pre-agricultural	Pre-maize
Pre-agricultural	Pre-maize
Pre-agricultural	Pre-maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Post-agricultural	Staple Maize
Pre-agricultural	Pre-maize
Post-agricultural	Staple Maize
Pre-agricultural	Pre-maize
Pre-agricultural	Pre-maize
Post-agricultural	Staple Maize

LMAMR_ID	Burial_ID	Age_14C	Dat	Site
MRSC-1888	ST.18.14.4	1335	1335	Saki Tzul
MRSC-1889	ST.17.5.4	1515	1515	Saki Tzul
MRSC-1890	ST.17.6.5	1590-1530	1560	Saki Tzul
MRSC-1891	ST.17.6.7a	1530	1530	Saki Tzul
MRSC-1892	ST.17.6.7b	1620	1620	Saki Tzul
MRSC-1893	ST.17.6.7c	1530	1530	Saki Tzul
MRSC-1894	MHCP.17.1.7	4625	4625	Mayahak Cab Pek
MRSC-1895	MHCP.14.2.4a	3990	3990	Mayahak Cab Pek
MRSC-1896	ST.16.1.3	6485	6485	Saki Tzul
MRSC-1897	ST.18.11.5	1235	1235	Saki Tzul
MRSC-1898	ST.18.14.12	2985	2985	Saki Tzul
MRSC-1905	MHCP.17.1.8	8565	8565	Mayahak Cab Pek
MRSC-1971	ST.18.11.8	4300	4300	Saki Tzul
MRSC-1973	ST.18.11.9	4500	4500	Saki Tzul
MRSC-2058	MHCP.98.19-20.misc	2600	2600	Mayahak Cab Pek
MRSC-2059	MHCP.98.19-20.misc2	2600	2600	Mayahak Cab Pek
MRSC-2060	ST.99.7.1	1800	1800	Saki Tzul
MRSC-2062	MHCP.14.1.P1	4500	4500	Mayahak Cab Pek
MRSC-2072	ST.99.1.P2	1800	1800	Saki Tzul
MRSC-2178	MHCP.19.13.10	~5,000-6,000	6000	Mayahak Cab Pek
MRSC-2296	MHCP.19.12.P12	~5,000	5000	Mayahak Cab Pek
MRSC-2304	MHCP.19.17.P2	~1,500	1500	Mayahak Cab Pek

Supplementary Table 1 Ancestral Maya metadata.

Sample	Source Study
ERR3307053	FellowsYates2021
ERR3579824	Velsko2019
ERR3579819	Velsko2019
PYK005-A	FellowsYates2021
CHS20	Fleskes2024
Abusir1519	Neukamm2020
MRSC-1895	This Study
ERR3307054	FellowsYates2021
Chimp_13C	Ozga2019
WAMP26	Honap2023
PYK003-A	FellowsYates2021
A18023-Edo-4	Eisenhofer2020
TO-2417Q-final	Bravo-Lopez2020
CS32	Velsko2019

CHS03	Fleskes2024
MRSC-1894	This Study
MRSC-2296	This Study
ATCC23263	NCBI - Reference
MAG	NCBI - ERR589356_bin.17_CONCOCT_v1.1_MAG
BIN1	NCBI

Supplementary Table 2 *Pseudoramibacter alactolyticus* comparative genomes.

Sample	**Source Study**
MRSC-1894	This Study
MRSC-1895	This Study
MRSC-1973	This Study
MRSC-1897	This Study
MRSC-2060	This Study
MRSC-2072	This Study
MRSC-1971	This Study
TO-2417Q	Bravo-Lopez2020
TO-2417J	Bravo-Lopez2020
WAMP03	Honap2023
WAMP29	Honap2023
WAMP13	Honap2023
WAMP22	Honap2023
WAMP33	Honap2023
WAMP23	Honap2023
UB4	NCBI
9601	NCBI
WW11663	NCBI
CHS31	Fleskes2024
92A2	NCBI
CS11	Velsko2019
CS15	Velsko2019
CS02	Velsko2019
CS32	Velsko2019
UB22	NCBI
ERR3307053	FellowsYates2021
ERR3579819	Velsko2019
ERR3307048	FellowsYates2021
UB20	NCBI
PYK002-A	FellowsYates2021

ATCC43037	NCBI
ERR3579823	Velsko2019
ERR3307054	FellowsYates2021
WW10960	NCBI
SRS019029	HMP2012
3313	NCBI
KS16	NCBI

Supplementary Table 3 *Tannerella forsythia* comparative genomes.

Sample	Total Reads	% Post Pre-Processing	Duplication Rate	AR Reads	Total % Retained	Average read length	Included in Analyses	Reason
EB030821	268201	99.50%	79.67%	37828	14.10%	83.0717	NA	BLANK
EB031221	108712	99.55%	93.67%	12590	11.58%	78.0259	NA	BLANK
EB081120	41868	99.65%	78.08%	12069	28.83%	84.6916	NA	BLANK
EB081220	112283	99.95%	94.20%	12573	11.20%	81.9756	NA	BLANK
EB081320	1056002	99.89%	95.99%	73859	6.99%	81.356	NA	BLANK
EB082119	360904	99.76%	22.71%	173545	48.09%	71.3201	NA	BLANK
EB121019	79998	99.96%	90.67%	12343	15.43%	81.1329	NA	BLANK
LB010623	1880937	99.24%	79.98%	440261	23.41%	141.938	NA	BLANK
LB041321	84234	99.20%	89.40%	14889	17.68%	74.7875	NA	BLANK
LB041921	451458	99.94%	90.88%	63331	14.03%	77.5368	NA	BLANK
LB082120	727903	99.71%	97.28%	38641	5.31%	77.59	NA	BLANK
LB082420	822108	99.88%	97.50%	42805	5.21%	80.0239	NA	BLANK
LB082620	718734	99.95%	95.83%	61132	8.51%	82.8429	NA	BLANK
LB090519	131534	99.15%	30.91%	65113	49.50%	67.5359	NA	BLANK
MRSC-1887	5738541	3.14%	8.86%	159581	2.78%	104.28	No	No oral signature
MRSC-1888	10643698	88.98%	1.08%	9367075	88.01%	82.8566	Yes	NA
MRSC-1889	4549887	89.65%	1.11%	4032310	88.62%	64.5303	Yes	NA
MRSC-1890	7002174	90.88%	1.21%	6285845	89.77%	78.4421	Yes	NA
MRSC-1891	9058551	89.24%	1.25%	7980222	88.10%	69.0875	Yes	NA
MRSC-1892	10376237	88.13%	1.03%	9047373	87.19%	68.0415	Yes	NA
MRSC-1893	5071761	86.36%	0.98%	4333679	85.45%	58.4689	Yes	NA
MRSC-1894	8346714	89.51%	1.76%	7331516	87.84%	58.4266	Yes	NA
MRSC-1895	75282291	89.62%	18.01%	55302318	73.46%	55.4012	Yes	NA
MRSC-1896	5683903	41.30%	8.64%	2142070	37.69%	63.2873	Yes	NA
MRSC-1897	7741132	87.58%	0.94%	6713717	86.73%	63.1158	Yes	NA

MRSC-1898	9409267	88.88%	1.22%	8260704	87.79%	64.1801	Yes	NA
MRSC-1905	1730233	71.13%	9.19%	1091901	63.11%	72.2081	Yes	NA
MRSC-1971	30262462	81.64%	14.44%	21084928	69.67%	61.9588	Yes	NA
MRSC-1973	17269295	74.05%	7.90%	11774210	68.18%	65.2606	Yes	NA
MRSC-1976	12952544	51.76%	15.32%	5617334	43.37%	102.991	No	No oral signature
MRSC-2057	21073739	54.16%	17.61%	9384795	44.53%	103.373	No	No oral signature
MRSC-2058	14284911	63.75%	10.22%	8167992	57.18%	81.5112	Yes	NA
MRSC-2059	2823718	56.38%	17.29%	1276616	45.21%	98.4177	Yes	NA
MRSC-2060	29660930	92.71%	9.69%	24824208	83.69%	77.1763	Yes	NA
MRSC-2061	33361140	81.46%	21.92%	21184324	63.50%	95.2402	No	No oral signature
MRSC-2062	19020251	82.42%	19.11%	12583555	66.16%	85.4356	Yes	NA
MRSC-2063	8886977	20.81%	17.09%	1425758	16.04%	107.252	No	No oral signature
MRSC-2072	38857045	94.82%	10.42%	33002470	84.93%	79.8998	Yes	NA
MRSC-2076	5415735	15.05%	21.10%	271083	5.01%	93.7616	No	Deduplicated reads < 500K
MRSC-2178	19646493	86.96%	25.45%	12660504	64.44%	69.98	Yes	NA
MRSC-2296	24535226	95.12%	19.44%	18780220	76.54%	70.301	Yes	NA
MRSC-2301	30378285	87.93%	19.67%	21440926	70.58%	99.152	No	No oral signature
MRSC-2304	3072615	60.05%	15.15%	882469	28.72%	104.808	Yes	NA
MRSC-2309	1756807	30.49%	16.25%	341861	19.46%	88.0463	No	Deduplicated reads < 500K
MRSC-2329	5978242	2.26%	20.33%	105149	1.76%	116.011	No	Not enough reads classified

Supplementary Table 4 Rock Shelter shotgun metagenome pre-processing summary

Reads retained post-processing includes trimmed and merged reads. AR reads are pre-processed, deduplicated, and filtered of human DNA sequences.

	Human Reference		Mitochondria	
Sample	**Unique mapped reads**	**% reference covered >= 1X**	**Unique mapped reads**	**% reference covered >= 1X**
MRSC-1888	1834	0%	2	0.68%
MRSC-1889	1021	0%	1	0.41%
MRSC-1890	898	0%	1	0.43%
MRSC-1891	2141	0%	3	0.99%
MRSC-1892	2798	0.01%	5	2.08%
MRSC-1893	3397	0.01%	2	1.51%
MRSC-1894	8045	0.03%	17	8.85%
MRSC-1895	13838	0.04%	163	54.87%
MRSC-1896	2316	0.01%	3	2.03%

MRSC-1897	2672	0%	2	0.50%
MRSC-1898	723	0%	0	0%
MRSC-1905	25294	0.08%	12	6.07%
MRSC-1971	53405	0.20%	103	53.32%
MRSC-1973	2950	0.01%	26	7.71%
MRSC-2058	7829	0.03%	5	4.10%
MRSC-2059	39162	0.20%	46	30.04%
MRSC-2060	7975	0.01%	18	7.47%
MRSC-2062	96041	0.39%	68	40.14%
MRSC-2072	5300	0.01%	7	2.21%
MRSC-2178	73141	0.31%	313	87.91%
MRSC-2296	21219	0.07%	12	8.40%
MRSC-2304	676611	2.53%	702	97.76%

Supplementary Table 5 Human reference and mitochondrial alignment summary.

AR reads were aligned to the T2T-CHM13v2.0 whole genome and rCRS mitochondrial reference genomes.

SUPPLEMENTARY MATERIAL B

Author List and Affiliations

Sarah J. Johnson[1,2], Tanvi P. Honap[1,2], Joy Li [1,2], Jacob J. Haffner[1,2], Alan Leventhal[4,5], Monica V. Arellano[5], Krithivasan Sankaranarayanan [1,6], Cara Monroe[1,2,3], Cecil M. Lewis, Jr.[1,2,3]

[1] Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, Norman, OK 73019, U.S.A.

[2] Department of Anthropology, University of Oklahoma, Norman, OK 73069, U.S.A.

[3] Center for the Ethics of Indigenous Genomic Research (CEIGR)

[4] Department of Anthropology, San Jose State University

[5] Muwekma Ohlone Tribe

[6] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, U.S.A.

Author Contributions

Conceptualization: C.M.L, T.P.H., C.M., K.S.

Resources: A.L., M.V.A.

Data curation: S.J.J., J.L., C.M.

Formal analysis: S.J.J.

Funding acquisition: C.M.L., T.P.H.

Investigation: S.J.J., J.L., J.J.H.

Methodology: S.J.J.

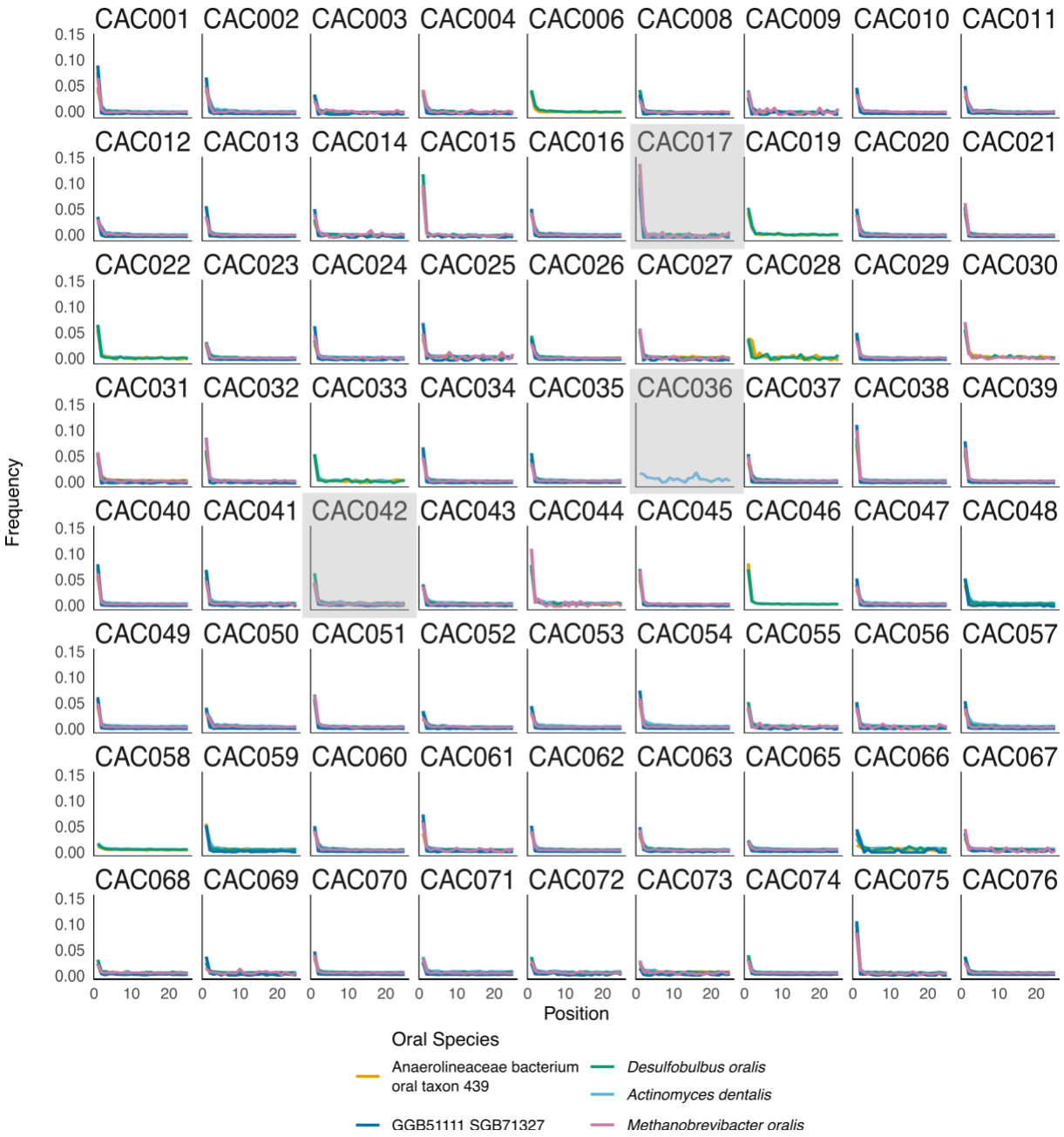
Supervision: C.M.L, T.P.H

Visualization: S.J.J.

Writing – original draft: S.J.J.

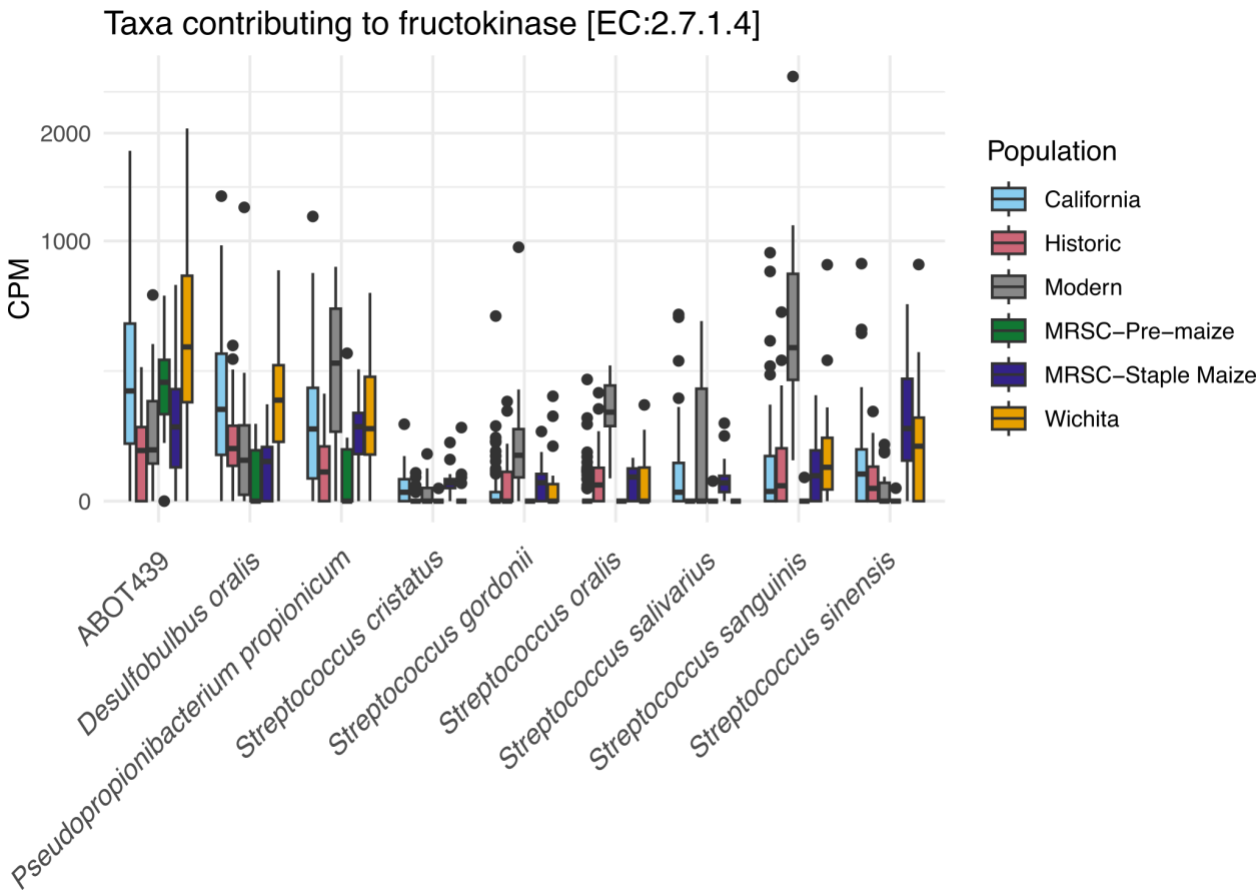
Writing – review and editing: All authors

Supplementary Figures



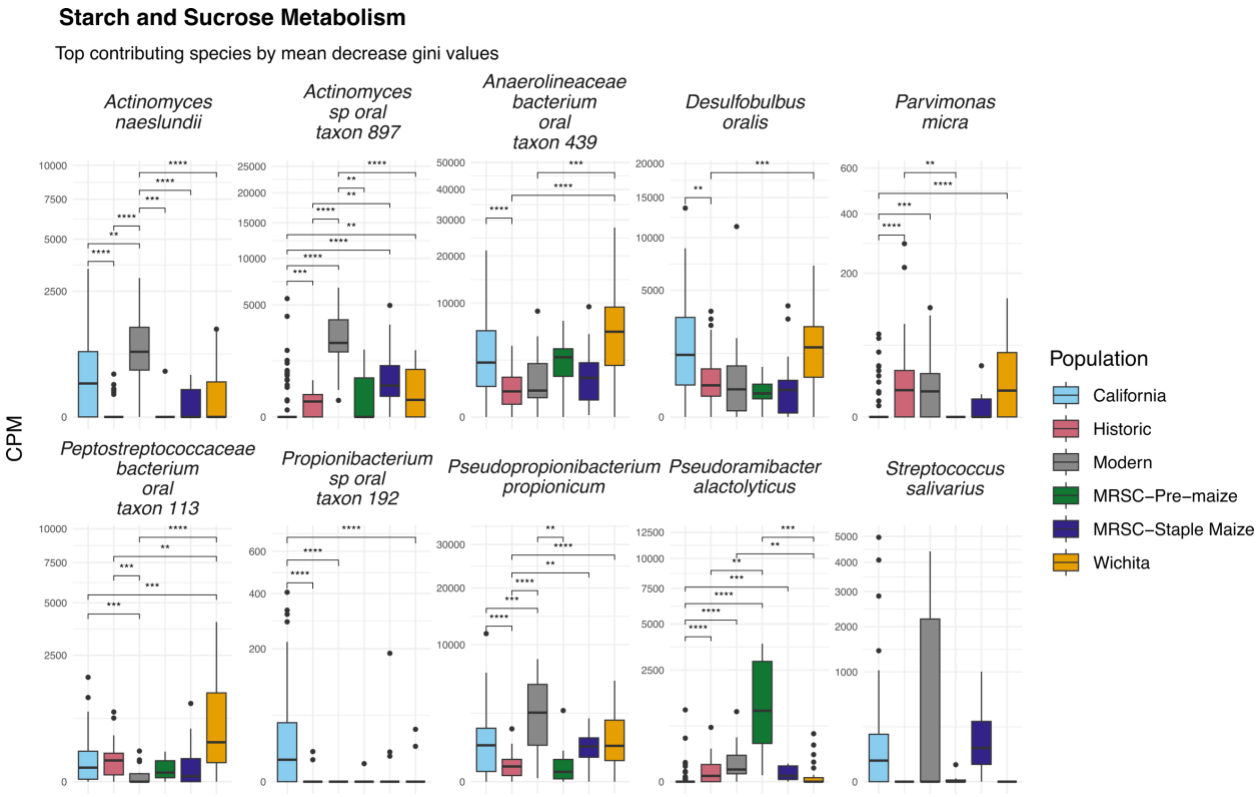
Supplementary Figure 5 Ancient damage plots from prevalent oral microbes.

Nucleotide misincorporations (C to T) at the 5` end of DNA reads of each sample. Frequencies determined by mapDamage2 (Jónsson et al., 2013) and plotted in R with ggplot2 (Wickham, 2016). Grey squares indicate samples already removed from downstream analysis due to poor oral preservation. CAC007 and CAC042 did not have enough aligned reads to form a damage plot.



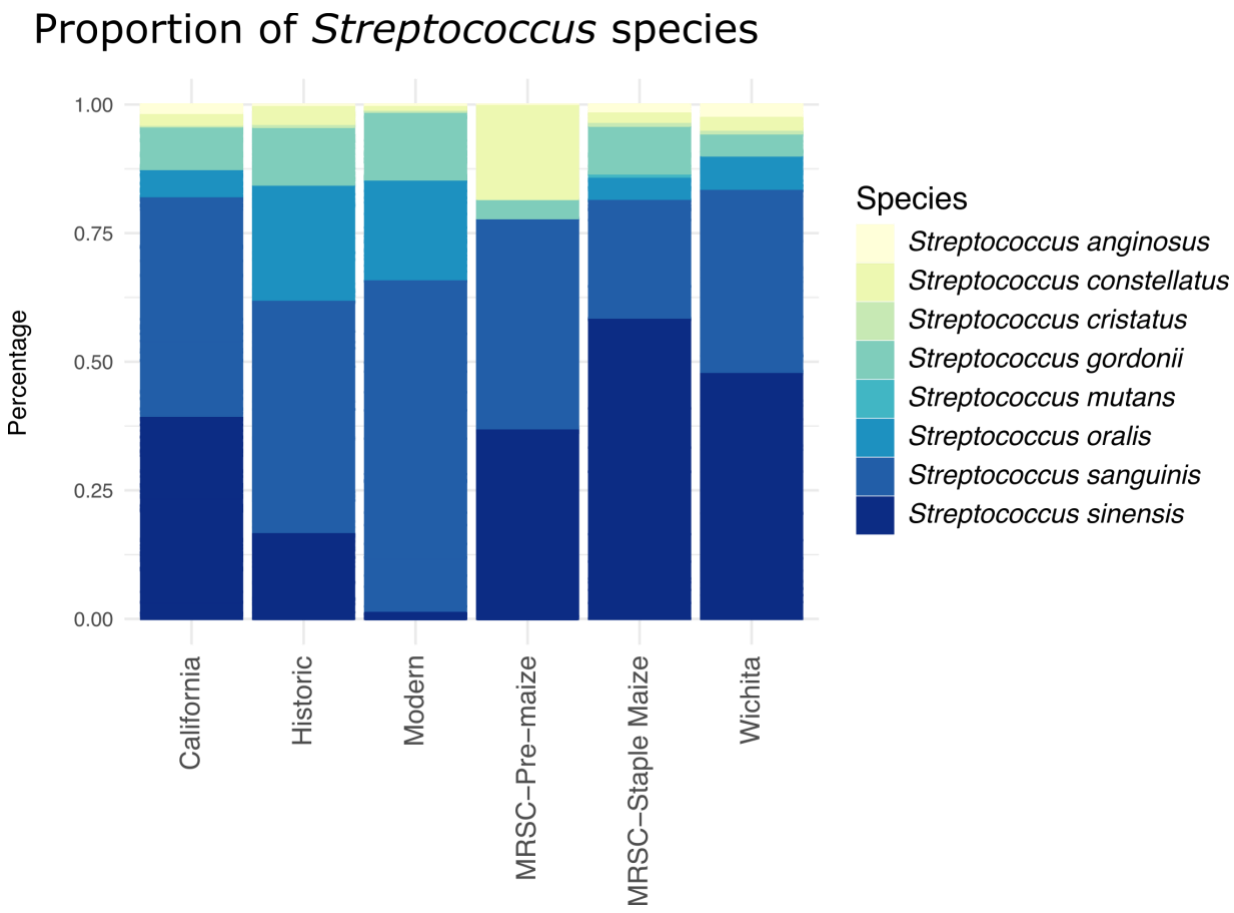
Supplementary Figure 6 Species contributing to Fructokinase

Species contribution to fructokinase function determined with HUMAnN3 (Beghini et al., 2021).



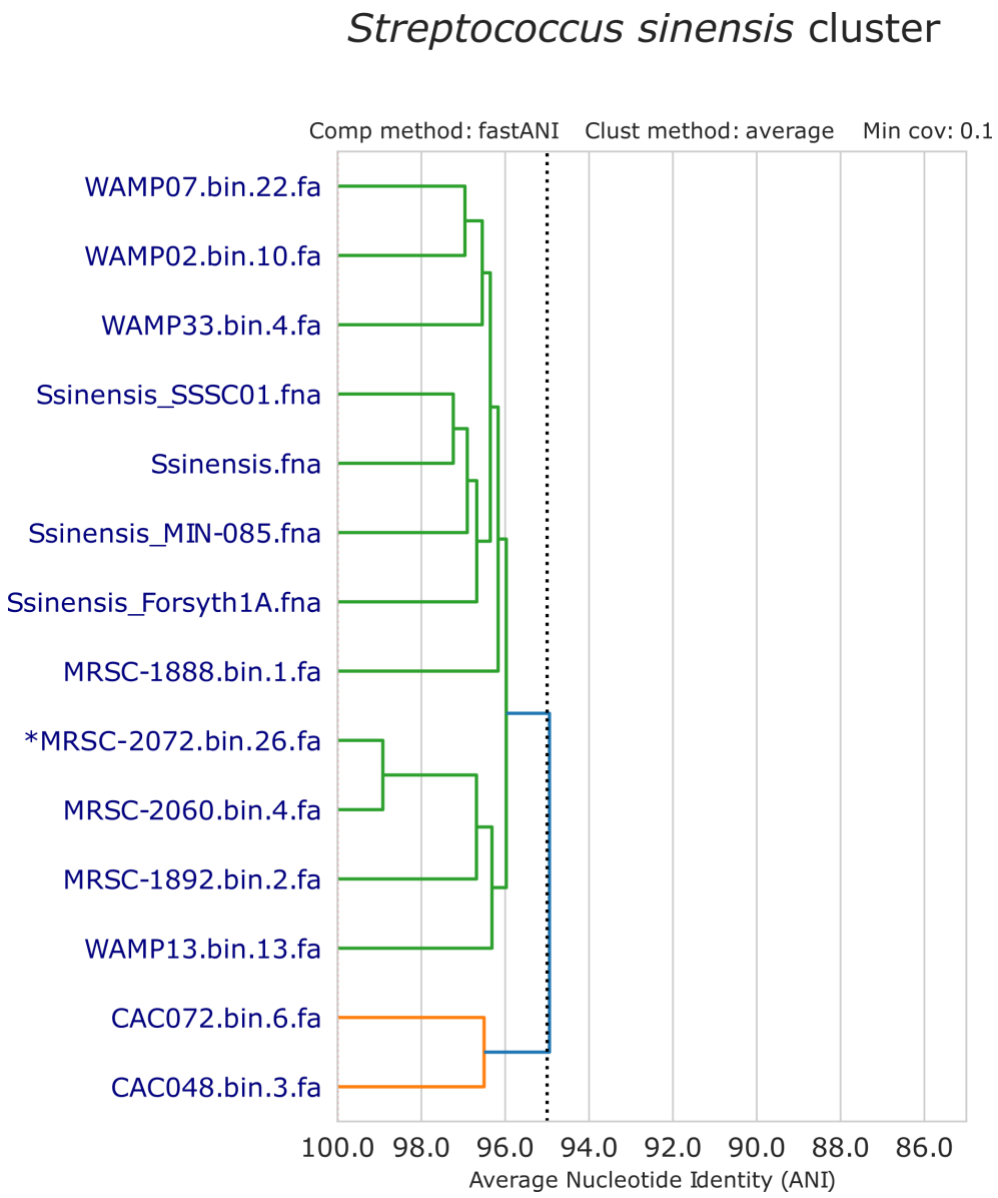
Supplementary Figure 7 Species contributing to starch and sucrose metabolism.

Stratified functional tables were subset to KO functions classified as “starch and sucrose metabolism”. Species were determined by a random forest model to have high mean decrease gini values in differentiating populations based on starch and sucrose metabolic function. The y-axis is scaled by square-root transformation for aesthetic purposes. Pairwise-Wilcoxon test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$).



Supplementary Figure 8 Proportion of *Streptococcus* species in each population.

Proportion of reads in each population classified by MetaPhlAn4 as one of depicted *Streptococcus* species.



Supplementary Figure 9 ANI clustering of *S. sinensis* MAGs and published genomes.

ANI clustering of *S. sinensis* MAGs and published genomes. ANI = average nucleotide identity.

Supplementary Tables

Sample	Total Reads	% Post Pre-Processing	Duplication rate	AR Reads	Total % Retained	Average Read Length	Included in Analysis	Reason

CAC001	33083058	94.2%	23.3%	23904803	72.3%	77.7	Yes	NA
CAC002	16145935	92.4%	18.9%	12092303	74.9%	67.6	Yes	NA
CAC003	9085366	92.6%	12.4%	7294036	80.3%	62.0	Yes	NA
CAC004	20822033	89.6%	17.2%	15441894	74.2%	65.5	Yes	NA
CAC005	2996469	87.7%	95.7%	113716	3.8%	65.4	No	No oral signature
CAC006	23781267	91.8%	61.5%	8378073	35.2%	73.2	Yes	NA
CAC007	403490	80.8%	15.5%	275327	68.2%	57.8	No	< 500,000 AR reads
CAC008	9691321	91.6%	5.7%	8368191	86.3%	72.6	Yes	NA
CAC009	49199405	66.3%	84.9%	4920756	10.0%	57.8	Yes	NA
CAC010	19966873	94.5%	18.8%	15318475	76.7%	68.5	Yes	NA
CAC011	22526014	91.4%	73.0%	5567766	24.7%	72.7	Yes	NA
CAC012	38353396	90.6%	21.8%	27132789	70.7%	71.5	Yes	NA
CAC013	59236453	94.4%	8.0%	51399482	86.8%	74.6	Yes	NA
CAC014	25694707	90.0%	59.0%	9474843	36.9%	86.3	Yes	NA
CAC015	655603	89.8%	6.0%	553011	84.4%	60.5	Yes	NA
CAC016	18055001	91.3%	11.6%	14548469	80.6%	80.7	Yes	NA
CAC017	5902090	93.5%	8.7%	5032554	85.3%	69.6	No	No oral signature
CAC018	15436800	75.1%	95.7%	498338	3.2%	69.3	No	No oral signature
CAC019	25147648	85.8%	66.7%	7172204	28.5%	60.9	Yes	NA
CAC020	21674616	90.7%	11.5%	17400003	80.3%	66.0	Yes	NA
CAC021	71015444	93.6%	10.3%	59610337	83.9%	68.6	Yes	NA
CAC022	18339126	90.4%	17.6%	13644570	74.4%	64.7	Yes	NA
CAC023	17140593	94.1%	9.9%	14519405	84.7%	71.5	Yes	NA
CAC024	30150182	87.5%	62.4%	9903397	32.8%	80.8	Yes	NA
CAC025	34453737	83.4%	94.4%	1595176	4.6%	75.0	Yes	NA
CAC026	16111337	94.1%	18.1%	12401208	77.0%	79.3	Yes	NA
CAC027	21945692	71.0%	9.9%	14038246	64.0%	64.8	Yes	NA
CAC028	13368234	48.0%	89.4%	678372	5.1%	53.2	Yes	NA
CAC029	37759287	94.3%	19.4%	28583387	75.7%	64.9	Yes	NA
CAC030	19948585	87.4%	11.6%	15413889	77.3%	61.5	Yes	NA
CAC031	15069353	93.3%	6.6%	13128135	87.1%	70.1	Yes	NA
CAC032	13290583	47.8%	32.6%	4279159	32.2%	56.8	Yes	NA
CAC033	43608562	3.9%	50.9%	839598	1.9%	55.0	Yes	NA
CAC034	23935835	91.8%	6.1%	20592064	86.0%	65.2	Yes	NA
CAC035	5733280	95.2%	8.9%	4967107	86.6%	69.9	Yes	NA
CAC036	6415869	92.2%	40.9%	3487402	54.4%	70.2	No	No oral signature
CAC037	80857142	91.5%	11.1%	65672330	81.2%	62.8	Yes	NA

CAC038	21678357	94.5%	11.4%	18134101	83.7%	66.4	Yes	NA
CAC039	24563903	92.0%	19.1%	18276298	74.4%	73.7	Yes	NA
CAC040	71919200	84.8%	43.2%	34540525	48.0%	57.8	Yes	NA
CAC041	28484967	88.9%	54.9%	11416321	40.1%	80.1	Yes	NA
CAC042	45754346	89.1%	74.0%	10602513	23.2%	80.0	No	No oral signature
CAC043	7107251	92.9%	11.7%	5810676	81.8%	59.6	Yes	NA
CAC044	12006139	77.4%	49.7%	4674570	38.9%	51.7	Yes	NA
CAC045	16049302	87.9%	65.2%	4892823	30.5%	61.6	Yes	NA
CAC046	67763024	75.4%	83.7%	8347527	12.3%	52.5	Yes	NA
CAC047	5840720	94.0%	9.5%	4966438	85.0%	66.6	Yes	NA
CAC048	16988328	91.2%	19.2%	12514790	73.7%	86.4	Yes	NA
CAC049	20547740	89.8%	20.4%	14669800	71.4%	69.1	Yes	NA
CAC050	24214254	94.0%	64.1%	8150580	33.7%	66.4	Yes	NA
CAC051	8567674	94.4%	12.5%	7069338	82.5%	64.5	Yes	NA
CAC052	2643226	90.4%	18.3%	1950848	73.8%	94.8	Yes	NA
CAC053	10909422	94.0%	6.3%	9604751	88.0%	74.4	Yes	NA
CAC054	82935021	93.3%	21.3%	60757104	73.3%	64.1	Yes	NA
CAC055	5956319	92.1%	9.8%	4946784	83.1%	62.5	Yes	NA
CAC056	20199876	90.7%	18.5%	14902948	73.8%	87.9	Yes	NA
CAC057	22996573	91.8%	10.1%	18974233	82.5%	68.5	Yes	NA
CAC058	10683360	92.2%	46.5%	5268478	49.3%	107.3	Yes	NA
CAC059	5909768	92.8%	13.0%	4769223	80.7%	61.3	Yes	NA
CAC060	17587140	87.4%	10.8%	13692510	77.9%	67.0	Yes	NA
CAC061	15893038	73.0%	25.3%	8650623	54.4%	65.1	Yes	NA
CAC062	94109461	94.5%	21.8%	69418848	73.8%	71.6	Yes	NA
CAC063	6698590	95.3%	10.4%	5691043	85.0%	65.1	Yes	NA
CAC064	12292925	2.3%	37.2%	177124	1.4%	68.1	No	Not enough reads classified
CAC065	30543864	94.7%	67.9%	9217021	30.2%	94.1	Yes	NA
CAC066	13728276	79.2%	90.1%	1069125	7.8%	96.8	Yes	NA
CAC067	45731717	82.2%	92.1%	2954557	6.5%	63.6	Yes	NA
CAC068	1188678	94.5%	1.3%	1106858	93.1%	87.4	Yes	NA
CAC069	1642920	95.2%	1.6%	1525603	92.9%	83.3	Yes	NA
CAC070	27172985	94.4%	4.3%	24533638	90.3%	84.0	Yes	NA
CAC071	2261848	94.9%	1.3%	2117293	93.6%	104.0	Yes	NA
CAC072	2162548	93.9%	1.3%	2002826	92.6%	109.1	Yes	NA
CAC073	1812535	92.0%	1.0%	1651243	91.1%	110.0	Yes	NA
CAC074	21356950	95.6%	2.9%	19824665	92.8%	98.0	Yes	NA
CAC075	2725271	95.4%	1.3%	2565648	94.1%	68.9	Yes	NA
CAC076	36994337	93.2%	2.3%	33653601	91.0%	105.8	Yes	NA

EB060122	4884805	86.5%	78.1%	923310	18.9%	66.8	NA	Blank
EB071122	1685635	81.9%	90.3%	132951	7.9%	69.1	NA	Blank
EB071222	1108130	77.6%	76.4%	202501	18.3%	70.0	NA	Blank
EB091322	1793287	0.9%	28.3%	10911	0.6%	64.0	NA	Blank
EB101722	1128097	31.2%	91.5%	24397	2.2%	71.0	NA	Blank
EB110722	4820302	88.2%	39.2%	2547962	52.9%	77.5	NA	Blank
LB060922	2798	31.7%	39.1%	501	17.9%	77.5	NA	Blank
LB072022	1015781	3.4%	86.1%	4813	0.5%	89.5	NA	Blank
LB092322	2025351	0.4%	55.4%	4023	0.2%	63.9	NA	Blank
LB102522	742107	4.3%	85.4%	4506	0.6%	58.4	NA	Blank
LB111722	1029241	3.8%	68.9%	11517	1.1%	85.5	NA	Blank
SCL-134 EC_calc_S12	766972	95.9%	22.6%	17816	2.3%	79.0	NA	Blank
SCL-134 EC1_S8	1289673	4.2%	47.1%	567628	44.0%	68.7	NA	Blank
SCL-134 LibNeg_S9	66580	68.2%	26.0%	33044	49.6%	87.9	NA	Blank

Supplementary Table 6 California shotgun metagenome pre-processing summary

Reads retained post-processing includes trimmed and merged reads. AR reads are pre-processed, deduplicated, and filtered of human DNA sequences.

SUPPLEMENTARY MATERIAL C

Author List and Affiliations

Sarah J. Johnson[1,2], Jacob J. Haffner[1,2], Tanvi P. Honap[1,2], Brandi Bethke[2,3], Leland C Bement[2,3], and Cecil M. Lewis, Jr.[1, 2]

[1] Laboratories of Molecular Anthropology and Microbiome Research (LMAMR), University of Oklahoma, Norman, OK 73019, U.S.A.

[2] Department of Anthropology, University of Oklahoma, Norman, OK 73069, U.S.A.

[3] Oklahoma Archaeological Survey, Norman, OK 73019

Author Contributions

Conceptualization: C.M.L, B.B., L.C.B., S.J.J.

Resources: B.B., L.C.B.

Data curation: S.J.J.

Formal analysis: S.J.J.

Funding acquisition: C.M.L., T.P.H.

Investigation: S.J.J., J.J.H.

Methodology: S.J.J.

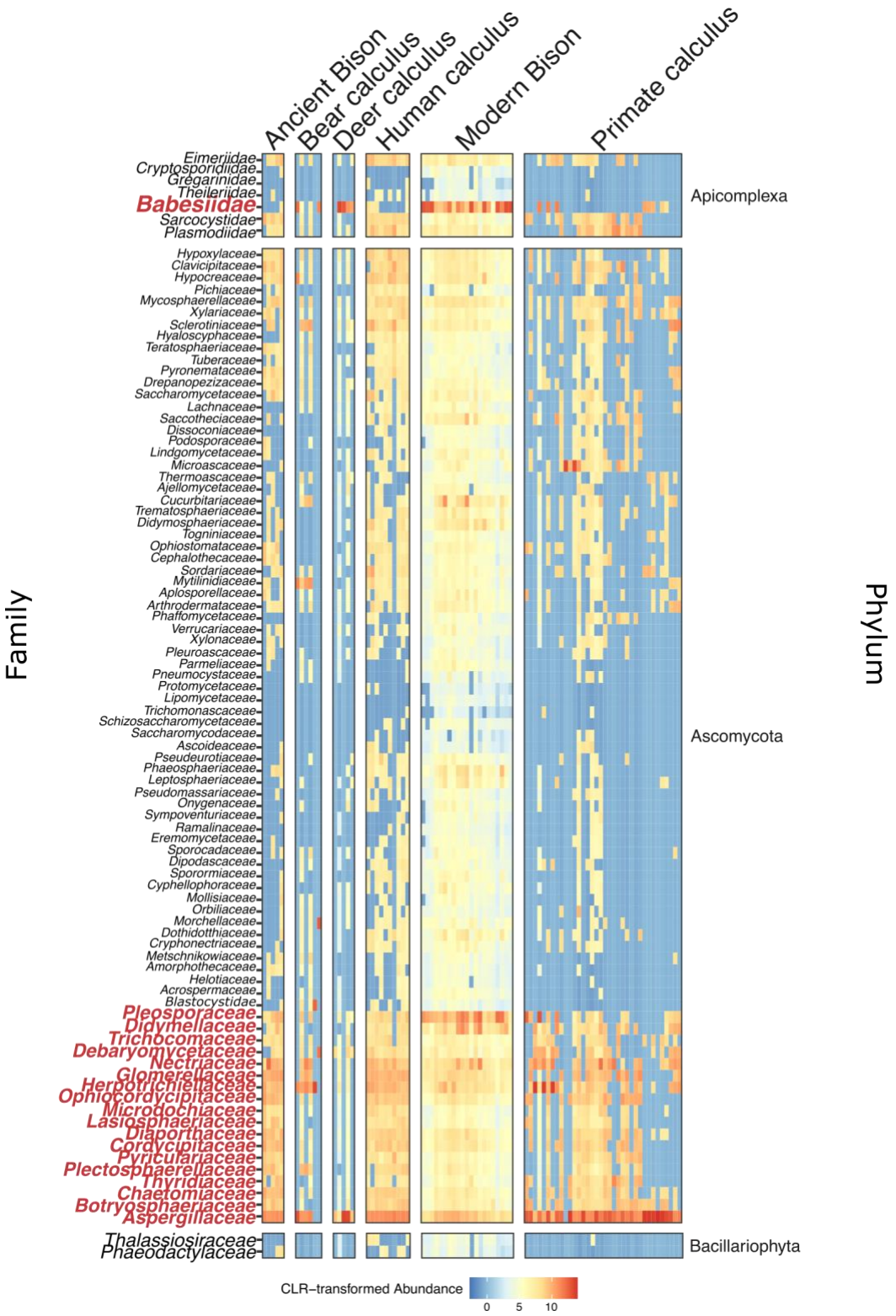
Supervision: C.M.L, T.P.H

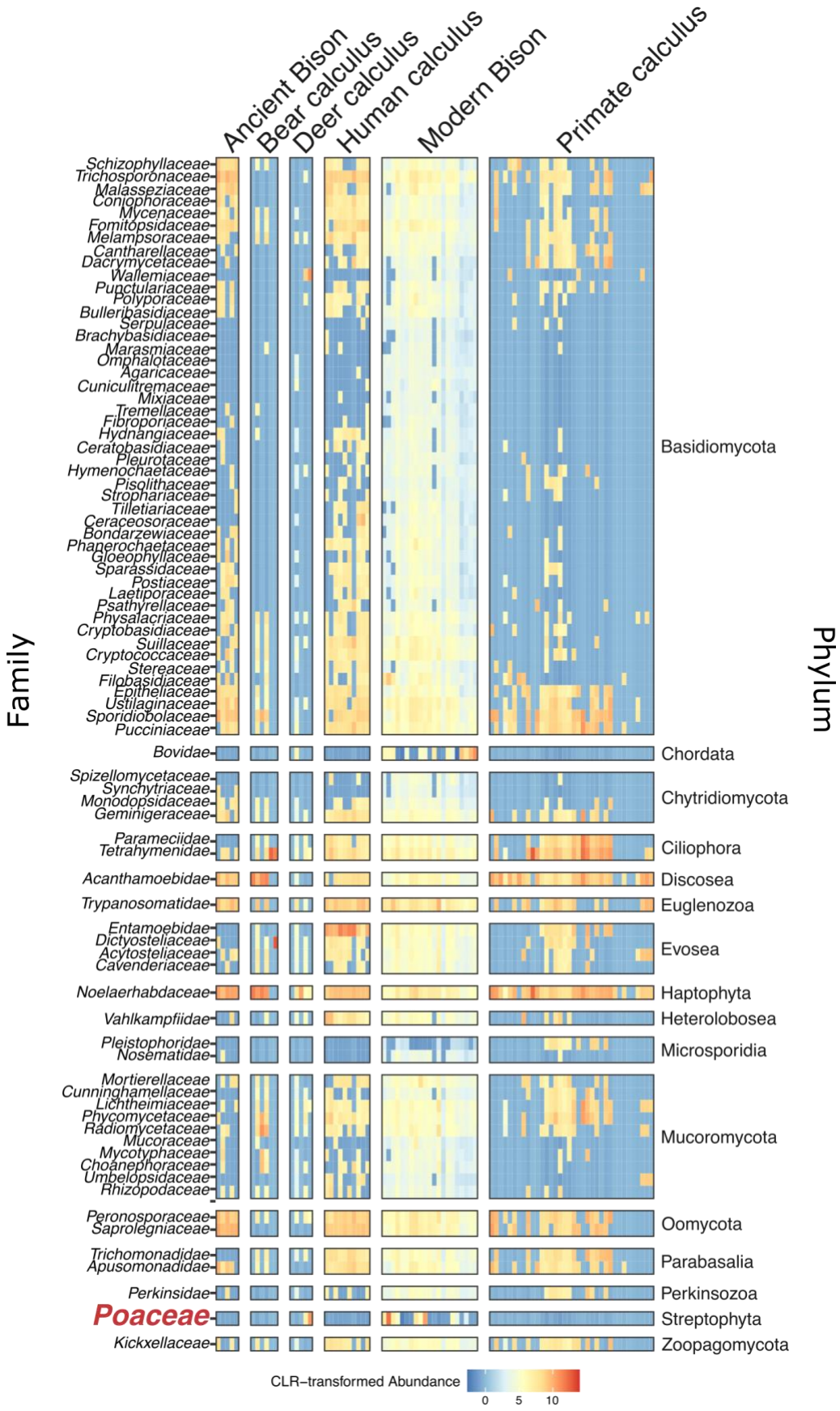
Visualization: S.J.J.

Writing – original draft: S.J.J.

Writing – review and editing: All authors

Supplementary Figures




Ancient Bison
Bear calculus
Deer calculus
Human calculus
Modern Bison
Primate calculus
Family
Phylum
Schizophyllaceae
Trichosporonaceae
Malasseziaceae
Coniophoraceae
Mycenaceae
Fomitopsidaceae
Melampsoraceae
Cantharellaceae
Dacrymycetaceae
Wallemiaceae
Punctulariaceae
Polyporaceae
Bulleribasidiaceae
Serpulaceae
Brachybasidiaceae
Marasmiaceae
Omphalotaceae
Agaricaceae
Cuniculitremaceae
Mixiaceae
Tremellaceae
Fibroporiaceae
Hydnangiaceae
Ceratobasidiaceae
Pleurotaceae
Hymenochaetaceae
Pisolithaceae
Strophariaceae
Tilletiariaceae
Ceraceosoraceae
Bondarzewiaceae
Phanerochaetaceae
Gloeophyllaceae
Sparassidaceae
Postiaceae
Laetiporaceae
Psathyrellaceae
Physalacriaceae
Cryptobasidiaceae
Suillaceae
Cryptococcaceae
Stereaceae
Filobasidiaceae
Epitheliaceae
Ustilaginaceae
Sporidiobolaceae
Pucciniaceae
Basidiomycota
Bovidae
Chordata
Spizellomycetaceae
Synchytriaceae
Monodopsidaceae
Geminigeraceae
Chytridiomycota
Parameciidae
Tetrahymenidae
Ciliophora
Acanthamoebidae
Discosea
Trypanosomatidae
Euglenozoa
Entamoebidae
Dictyosteliaceae
Acytosteliaceae
Cavenderiaceae
Evosea
Noelaerhabdaceae
Haptophyta
Vahlkampfiidae
Heterolobosea
Pleistophoridae
Nosematidae
Microsporidia
Mortierellaceae
Cunninghamellaceae
Lichtheimiaceae
Phycomycetaceae
Radiomycetaceae
Mucoraceae
Mycotyphaceae
Choanephoraceae
Umbelopsidaceae
Rhizopodaceae
Mucoromycota
Peronosporaceae
Saprolegniaceae
Oomycota
Trichomonadidae
Apusomonadidae
Parabasalia
Perkinsidae
Perkinsozoa
Poaceae
Streptophyta
Kickxellaceae
Zoopagomycota
CLR−transformed Abundance
0
5
10

Supplementary Figure 10 Heatmap of eukaryotic taxa.

Includes taxa at the family level present in at least 25% of samples per group. Families ordered by hierarchical clustering of Euclidean distance.

Supplementary Tables

Sample	Total Reads	% Post pre-processing	Duplication Rate	% Retained After Human Filtering	Bison filtered reads	Total % Retained	Average length of filtered reads	Project
BIS001	18785229	90.7%	12.9%	79.1%	14849503	79.0%	60.8	Cooper Site Archaeological Bison
BIS002	15478280	90.2%	79.7%	18.3%	2832616	18.3%	73.9	Cooper Site Archaeological Bison
BIS003	16944049	87.9%	75.5%	21.5%	3645841	21.5%	67.7	Cooper Site Archaeological Bison
BIS004	18307836	91.6%	82.9%	15.7%	2873780	15.7%	69.5	Cooper Site Archaeological Bison
BIS005	17261988	92.9%	84.7%	14.2%	2444000	14.2%	84.3	Cooper Site Archaeological Bison
CWB01A	114602692	100.0%	7.9%	92.1%	18398208	16.1%	152.2	Clear Water Bison Ranch
CWB01B	105677559	100.0%	7.1%	92.8%	17720980	16.8%	147.0	Clear Water Bison Ranch
CWB01C	81274188	100.0%	11.9%	88.1%	18134886	22.3%	150.9	Clear Water Bison Ranch
CWB02	100368555	100.0%	9.8%	90.2%	85457379	85.1%	158.0	Clear Water Bison Ranch
CWB03A	76493446	99.9%	10.2%	89.7%	53921058	70.5%	165.7	Clear Water Bison Ranch
CWB03B	92907607	100.0%	7.0%	93.0%	62741048	67.5%	153.2	Clear Water Bison Ranch
CWB04A	105294654	100.0%	15.7%	84.3%	75440961	71.6%	131.4	Clear Water Bison Ranch
CWB04B	86270816	100.0%	15.3%	84.7%	72057288	83.5%	152.0	Clear Water Bison Ranch
CWB05A	95686623	100.0%	7.4%	92.5%	61424559	64.2%	150.9	Clear Water Bison Ranch

CWB05B	98193118	100.0%	7.3%	92.7%	78695898	80.1%	157.5	Clear Water Bison Ranch
CWB06	81373287	100.0%	8.2%	91.8%	71335259	87.7%	162.2	Clear Water Bison Ranch
CWB07A	70187977	100.0%	8.0%	92.0%	18694777	26.6%	139.4	Clear Water Bison Ranch
CWB07B	81309299	99.9%	19.7%	80.3%	56136716	69.0%	153.1	Clear Water Bison Ranch
CWB08	61557980	99.9%	17.1%	82.8%	40626508	66.0%	171.8	Clear Water Bison Ranch
CWB09A	84143308	100.0%	7.0%	92.9%	17463553	20.8%	160.3	Clear Water Bison Ranch
CWB09B	98366059	99.9%	15.6%	84.3%	54387445	55.3%	159.2	Clear Water Bison Ranch
CWB09C	110582618	99.9%	8.2%	91.8%	101061963	91.4%	157.9	Clear Water Bison Ranch
CWB10	92864558	100.0%	7.6%	92.4%	25369299	27.3%	137.5	Clear Water Bison Ranch
CWB11A	110899144	99.9%	12.3%	87.7%	71380070	64.4%	159.9	Clear Water Bison Ranch
CWB12A	36980286	99.9%	10.8%	89.2%	22683173	61.3%	145.3	Clear Water Bison Ranch
CWB12B	64155136	99.9%	14.0%	85.9%	34125645	53.2%	167.8	Clear Water Bison Ranch
EB030424	137176	99.5%	9.9%	89.7%	115674	84.3%	129.7	Clear Water Bison Ranch
EB041024	156218	99.6%	20.6%	79.1%	122366	78.3%	124.3	Clear Water Bison Ranch
EB041724	127060	98.5%	20.9%	77.9%	96547	76.0%	149.7	Clear Water Bison Ranch
EB081922	1951522	32.6%	88.0%	3.9%	73223	3.8%	86.0	Cooper Site Archaeological Bison
EB092923	2099984	99.9%	7.1%	92.8%	1872064	89.1%	131.4	Clear Water Bison Ranch
LB051424	28320	98.2%	27.0%	71.7%	19575	69.1%	148.6	Clear Water Bison Ranch
LB082922	1382657	1.5%	87.3%	0.2%	2620	0.2%	104.7	Cooper Site Archaeological Bison

Supplementary Table 7 Bison shotgun metagenome pre-processing summary

Reads retained post-processing includes trimmed and merged reads. Percent retained after removing human reads is displayed. Reads with Bison sequences removed were used for taxonomic classification and alignments.